Does planning explain why predictability affects reference production?

Sandra A. Zerkle   SZERKLE@UNC.EDU
Department of Psychology & Neuroscience, University of North Carolina – Chapel Hill

Jennifer E. Arnold   JARNOLD@EMAIL.UNC.EDU
Department of Psychology & Neuroscience, University of North Carolina – Chapel Hill

Abstract

How does thematic role predictability affect reference production? This study tests a planning facilitation hypothesis – that the predictability effect on reference form can be explained in terms of the time course of utterance planning. In a discourse production task, participants viewed two sequential event pictures, listened to a description of the first picture (depicting a transfer event between two characters), and then provided a description of the second picture (continuing with one thematic role character, either goal or source). We replicated previous findings that goal continuations lead to more reduced forms of reference and shorter latency to begin speaking than source continuations. Additionally, we tracked speakers' eye movements in two periods of utterance planning, early vs. late. We found that 1) early planning supports the use of reduced forms but is not affected by thematic role; 2) thematic role only affects late planning; and 3) in contrast with our hypothesis, planning does not account for predictability effects on reduced forms. We then speculate that discourse connectedness drives the thematic role predictability effect on reference form choice.

Keywords: utterance planning, thematic role predictability, reference production, discourse

1   Introduction

Variation in reference form exists in natural language production: an entity could be referred to with its proper name (*Lady Mannerly*), a noun phrase description (*the duchess*), a reduced form such as a pronoun (*she*), or even dropping the referent entirely (*…and Ø left*). What guides speakers as they make this choice? Previous work has supported the hypothesis that certain discourse conditions promote the use of particular referential forms (e.g., Ariel, 1990, 2001; Chafe, 1976, 1994; Gundel, Hedberg, & Zacharaski, 1993), where reduced forms tend to be used for salient or accessible referents. For example, speakers strongly prefer pronouns to refer to entities that appeared recently in the discourse, and especially in the previous subject position (Arnold, 1998). In addition, in some cases speakers are also more likely to use pronouns for thematic roles that are predictable (Arnold, 2001; Rosa & Arnold, 2017; but see Fukumura & van Gompel, 2010; Rohde & Kehler, 2014). However, there are no explicit models of the production mechanisms underlying reference production.

In particular, an unsolved puzzle concerns the role of predictability in pronoun production, and how and why predictability affects referential choices. The kind of predictability that matters here is **referential predictability**, i.e. the likelihood that the speaker will mention that entity at that point in the discourse. This is related to the calculation of **referential probability,** which includes probability estimates made both before and after the referring expression.

There is good evidence that predictability is important for comprehension, for example listeners are faster to understand predictable information, and sometimes even anticipate it (e.g., Altmann & Kamide, 1999). However, it's not obvious what predictability would mean to the speaker. People usually know what they are going to say ahead of time, so they aren't predicting their message; they are planning it. That is, while **predictability** can be thought of as a property

of the referent within context, it is not known whether this leads to actual **prediction** in all cases. This raises questions about how factors related to predictability in comprehension might affect production, and why.

Despite the oddness of prediction during production, it is important to investigate the relationship between predictability and reference production, because referential probability underlies numerous models of reference production (Frank & Goodman, 2012; Tily & Piantadosi, 2009), other aspects of language production (Hale, 2001; Levy & Jaeger, 2007), and reference comprehension (Frank and Goodman, 2012; Kehler & Rohde, 2013). However, empirical findings show that referential probability only sometimes affects pronoun use (see e.g. Rosa & Arnold, 2017 on the one hand, and e.g. Rohde & Kehler, 2014 on the other), while it has a stronger effect on prosodic variation (e.g., Arnold & Watson, 2015). This raises questions about how predictability really affects reference form. Here we address this issue by examining the relation between reference form, referential predictability, and utterance planning.

1.1   Thematic role predictability affects reference form

Previous work has revealed that reduced forms of reference are more likely in certain linguistic contexts: if an entity is given (Fowler & Hossum, 1987), in the subject position (Arnold, 1998; Chafe, 1976), contextually salient, topical, in focus, accessible (Ariel, 1990; Arnold, 2010; Chafe, 1994; Frank & Goodman, 2012; Givón, 1983; Grosz et al., 1995), or predictable in context (Jurafsky, Bell, Gregory, & Raymond, 2001; Levy & Jaeger, 2007; Mahowald et al., 2013). These findings span multiple types of reduction, including durational shortening, the use of reduced syntactic structures, as well as the use of pronouns instead of longer referential expressions. The current study is focused on the contrast between explicit descriptions (*the duchess)* and pronouns or dropped (zero) references.

We focus our attention on how thematic roles influence referential predictability, and whether this affects reference form. Thematic roles are assigned to a verb's arguments (Dowty, 1991; Jackendoff, 1972), and represent a semantic characterization of the entities involved in an event. The current study examines the thematic roles associated with transfer verbs, where the verb of a sentence depicts the transfer of an object (the theme) from one character (the source) to another character (the goal), as in "*the duchess handed a painting to the duke*". In a discourse context, these thematic role positions affect people's expectations about who will be mentioned next, where the goal referent is more expected than the source (Rosa & Arnold, 2017; Stevenson et al., 1994). Rosa and Arnold measured this expectation by simply asking participants who they thought would be mentioned next, but it also correlates with a tendency for speakers to mention the goal more in either natural discourse (Arnold, 2001) or story-continuation tasks (Stevenson et al., 1994).

There are a few possibilities for why goals are frequently mentioned, and perceived as predictable referents. One idea is that people expect goals because they frequently hear other people mention goals, so this high frequency affects future expectations of similar events (Arnold, 1998, 2001; Rosa & Arnold, 2017). Another possibility is that people have an inherent interest in goals, such that we expect the next event to involve the recipient of a transferred object.

When we consider transfer verbs, there is also excellent evidence that thematic roles affect reference form. Speakers are more likely to use pronouns and zeros when referring to goals than sources (Arnold, 2001; Rosa & Arnold, 2017; Zerkle, Rosa, & Arnold, 2015). Some of this evidence comes from a paradigm (also used in the current study) in which participants saw pairs of pictures depicting a transfer in the first picture and a subsequent action in the second picture (see Figure 1). For half of the items, the goal character continued in the subsequent event, and for

the other half the source character continued. Participants listened to a description of the first picture, and were instructed to provide a description of the second picture. In three experiments (Rosa & Arnold, 2017, exp. 1; Zerkle et al. 2015, exps. 1 and 2), speakers used reduced forms more for goal than source continuations. This finding occurred on top of the well-known tendency for subject continuations to use a pronoun. See Figure 2 for reference form choice results from these three studies.

Figure 1. Example trial: A speaker listens to a description of the first panel ("*The duchess handed a painting to the duke*") and then must provide a description of the second panel (e.g., "*and then he threw it in the closet*").
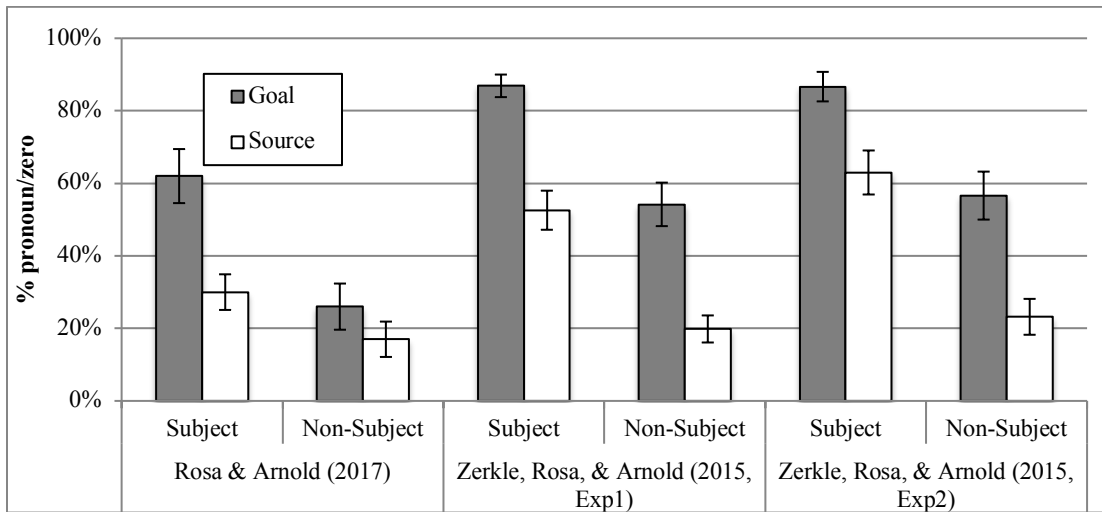

Figure 2. Speakers tend to use reduced forms more often for reference to goal continuations than source continuations, as well as for reference to subject continuations than non-subject continuations. (Note that Rosa & Arnold (2017) findings reflect pronoun use only, whereas the reduced forms in Zerkle, Rosa, & Arnold (2015) include both pronouns and zeros). Error bars are standard error of the subject means by condition.

One interpretation of these findings is that they reflect a preference to use reduced expressions for predictable information (Arnold, 2001; Rosa & Arnold, 2017). This is consistent with other evidence that speakers tend to use reduced forms for redundant information (Levy & Jaeger, 2007; Mahowald et al., 2013), and in particular that speakers use pronouns more for

predictable referents (Arnold, 1998, 2010). However, this conclusion is complicated by the fact that not all types of predictability affect reference form. Several studies have examined implicit causality contexts, such as *Val admired Elyce because....* or *Val impressed Elyce because….* Listeners have a strong expectation that the speaker will mention the person who is the more likely cause of the event (i.e., *Elyce* in the first sentence; *Val* in the second). However, several studies have found that this expectation has no effect on the speakers' use of pronouns (Fukumura & van Gompel, 2010; Kehler et al., 2008; Kehler & Rohde, 2013; Kravtchenko et al. 2017, but see Weatherford & Arnold, under review). Instead, speakers use pronouns more for the subject (*Val*) than the object (*Elyce*).

While this difference between transfer verb experiments and implicit causality experiments may be attributable to the different verbtypes, it raises questions about why the thematic role effect occurs. In addition, it's not clear how thematic role predictability affects language production mechanisms. We assume that in most cases, speakers do not need to predict their own speech, because typically conceptual planning precedes formulation (for an alternate view see Pickering and Garrod, 2013). Instead, we expect that the types of conditions that make referents predictable to the listener may be associated with the production process.

Here we test the hypothesis that the tendency to use reduced expressions for goals is not a direct result of prediction per se, but rather is a consequence of the mechanisms involved in utterance planning. Perhaps goal continuations are easier to plan (conceptually and/or lexically) than source continuations, and it is this facilitation that drives the choice to use a reduced form. The referential predictability previously found for goal continuations could affect both the time course of planning and reference form choice, raising the question of whether or not planning has a direct effect on reference form.

1.2   The role of planning facilitation

Hypothetically, there are several reasons why ease of planning might affect reference form. One possibility is that early planning might increase the salience of information in the speaker's discourse model. Classic theories about reference production suggest that pronouns are selected when the referent is salient, accessible, or in focus (for a review see Arnold, 2016; Arnold & Zerkle, 2019), where referential salience is often modeled in terms of the activation of the referent's conceptual representation (Arnold & Griffin, 2007; van Rij et al., 2010). When planning is easy, it may leave greater mental resources for representing information in the discourse model, increasing the activation on those referents. Thus, if predictability makes planning easy, this might explain the effect of predictability on reference production. Another possibility, suggested by rational models of reference form, is that the production of reduced forms incurs a lower cognitive cost than the production of explicit forms (e.g., Frank & Goodman, 2012; Tily & Piantadosi, 2009).

One of the few studies to address the relation between utterance planning and reference form is Arnold & Nozari (2017). In their study, speakers described shapes moving on screen, e.g. *The yellow pentagon flashes. Then it jumps over the yellow square.* They found that for the sham subjects, who were not experiencing anodal transcranial direct current stimulation, pronouns and zeros were used more often when the timing of the stimuli and response enabled the speaker to do more pre-planning, i.e., when there was no gap between utterances, and when the speaker was planning one utterance while describing the previous event. In this task, these conditions also promoted other measures of discourse connectivity, such as the use of connectors *and* or *then*. However, this study used discourse contexts with relatively little semantic information, and did not address questions of how planning relates to predictability.

The current study uses an eyetracking method to test whether planning facilitation explains the role of predictability on reference production. On the planning facilitation hypothesis, thematic roles promote the status of predictability, which in turn triggers earlier conceptual activation of the referent, making it more accessible for faster planning, which leads to increased pronoun choice.

In sum, the main question we will be asking in the current study is whether or not planning mechanisms can explain the predictability effect on reference form choice, and we will be using fine-grained measures of utterance planning with eye movements to test this. We will examine anticipatory looks to the target panel in two broad regions, representing relatively early and late planning windows.

1.3   Traditional models of utterance planning

Traditional models of utterance planning generally agree that the generation of an utterance is incremental; beginning with the transformation of a communicative intention into a preverbal conceptual message. The speaker then encodes the grammatical formulation of this message (which includes lexical/lemma selection and functional assignment), phonological encoding and syllabification occurs, and then the phonemes are encoded and articulated (Bock, Irwin, Davidson, & Levelt, 2002; Ferreira & Swets, 2002; Griffin & Bock, 2000; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999; Wheeldon, 2013; see Figure 3). Visual input can affect the early conceptual message level, such as in picture naming tasks (Indefrey & Levelt, 2004; Roelofs, 1992).
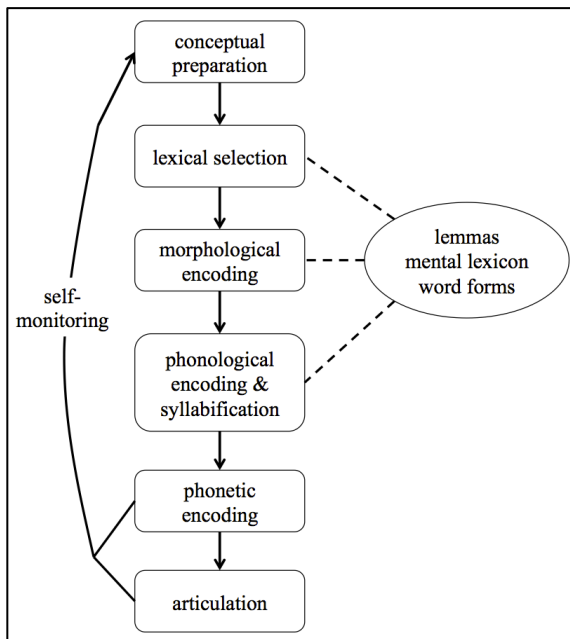

Figure 3. Model of utterance planning recreated from Levelt, Roelofs, & Meyer (1999).

Following the general model in Figure 3, Schmitt, Meyer, & Levelt (1999) built a model of planning that is specific to reference production. They proposed that speakers must fit each message fragment into the current discourse, which involves marking certain concepts as old

information (in focus), and others as new (not in focus; Chafe, 1976).[1] These types of information can be referred to linguistically in different ways, from explicit noun phrases to reduced pronouns. Such marking helps speakers and listeners maintain a coherent discourse record (Levelt, 1989). In order to access pronouns, lexical concepts of the entities are marked as being in focus or not in focus depending on the discourse context. A pronoun is selected if and only if the corresponding lexical concept is accessible and marked as in focus within the discourse; otherwise, a noun is selected. The authors propose that a referent's discourse accessibility status affects the message at the conceptual level; this accessibility node then affects the anaphoric link to lexical selection, in this case whether a reduced form vs. an explicit form is chosen (see Figure 4).

However, this model did not address questions about the timing of utterance planning, and whether planning is faster for accessible referents. Neither does it include consideration of predictability as a component of discourse status. This raises questions about whether planning facilitation has any effect at all on reference production.
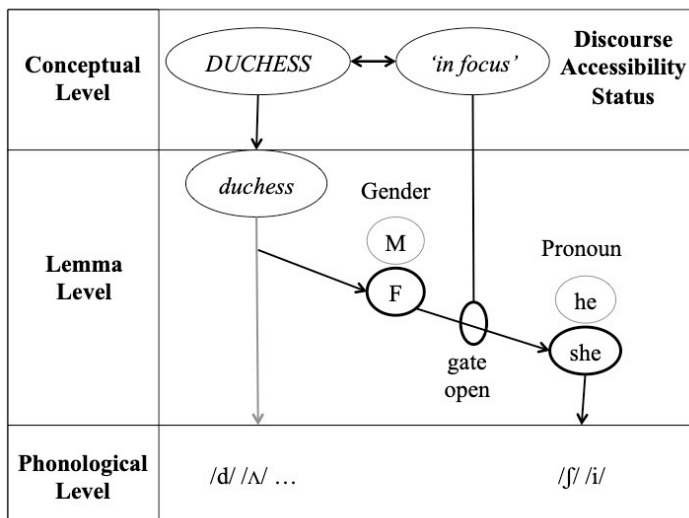

Figure 4. An English-language adaptation of the model of reference production from Schmitt, Meyer, & Levelt (1999).

1.4   Does planning explain thematic role effects on reference form?

The question behind our study is whether planning explains the relation between thematic roles and reference production choices. We have preliminary evidence about this question from previous studies (Rosa & Arnold, 2017; Zerkle, Rosa & Arnold, 2015), which used response latency as a proxy measure of planning time. In the storytelling task described above, participants hear the first sentence, and then provide a second sentence by describing the second panel of the cartoon picture. The time to respond provides a rough measure of the time needed to plan the response. Some planning could take place earlier, as participants listened to the first sentence. However, there was variable time to respond across trials, suggesting that considerable planning occurred immediately before the utterance.

---

[1] Note that on this terminology, "in focus" refers to information that is salient or accessible. This contrasts with the linguistic term focus, which contrasts with topic and is associated with new information.

In all three experiments, they found that speakers initiated goal continuations faster than source continuations, supporting the hypothesis that predictability affects the time course of utterance planning (see Figure 5). Response latency was measured from the offset of the initially heard description to the onset of the speaker's fluent speech. [2]
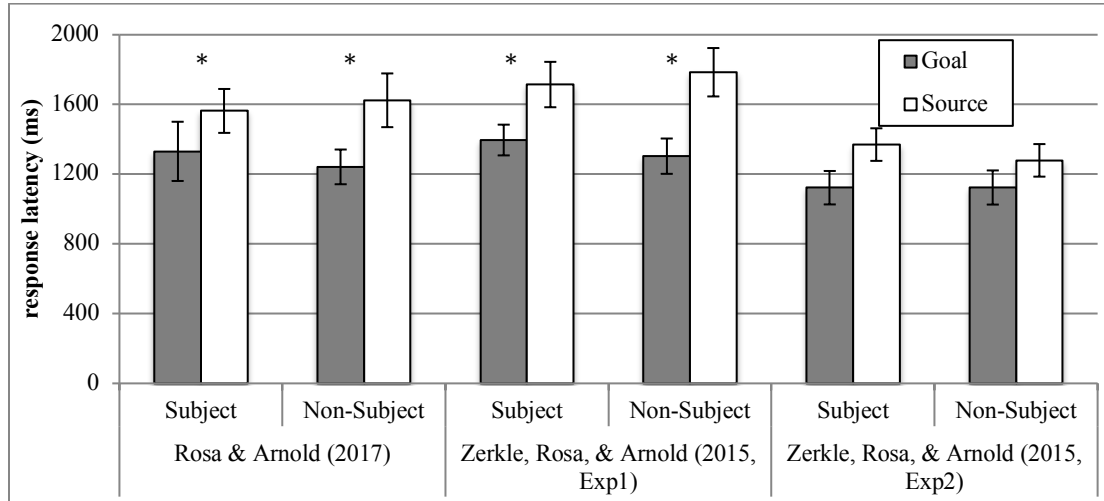

Figure 5. Main effect (indicated by asterisks) of goal continuation on response latency in Rosa & Arnold (2017) and Zerkle, Rosa, & Arnold (2015, Exp. 1); numerical trend in Zerkle, Rosa, & Arnold (2015, Exp. 2). Error bars are standard error of the subject means by condition.

In sum, goal continuations both yield faster response latencies and are more likely to involve pronouns. The next question is whether response latency itself affects reference form choice. Yet all three of these experiments found that latency does not predict reference form – as latency to begin speaking increased, there was no significant change in the rate of pronoun/zero choice in any of these three experiments (analyzed separately).

These findings provide strong evidence against one sort of hypothesis about planning time. However, planning in a discourse context can often occur earlier, while hearing/producing earlier utterances. For example, in Rosa & Arnold (2017, Exp. 1), participants could see both panels throughout the trial, allowing earlier planning. In both of the Zerkle et al. (2015) experiments, participants did not see the second panel until after the first sentence, but they had previewed the stimulus stories before the production task, which means they may have recalled the target sentence from memory before it appeared. Both of these tasks mirror the conditions of real life production, where the speaker typically knows what they are going to say before they say it. The current study tests whether earlier measures of planning pattern with reference form choices, and whether this explains the effects of predictability.

1.5 Current study goals
The current study tests the relation between utterance planning, predictability, and reference form. Even though latency did not predict reference form choice, eyetracking may provide an earlier and more sensitive measure for identifying planning effects on reference production. The

---

[2] In Rosa & Arnold (2017), this effect of predictability on latency occurred even after controlling for the length of the context sentence (Arnold, 2017), which may have affected participants' ability to pre-plan their responses. In the experiments from Zerkle, Rosa & Arnold (2015), the stimulus picture did not appear until after the context sentence, so the duration of the context sentence is not as relevant. See Zerkle, Rosa, & Arnold (2017) for a discussion of differences in planning across these two studies.

conceptual planning of a message could begin earlier than the latency period that was used in previous studies. Indeed, speakers often pre-plan their utterances (e.g, Ferreira & Swets, 2002), at least at the message level (Brown-Schmidt & Konopka, 2015). If speaking about predictable entities leads to greater pre-planning than speaking about unpredictable entities, it might explain the link between predictability and reference production.

To address this, in the current study we used eye tracking to each visual scene as a proxy for planning. Previous research has suggested that fixations to visual stimuli reflect the conceptual and linguistic planning that occurs before utterance production. In a simple picture-description task, trial-initial looks signify rapid planning: speakers may direct their attention to the part of the displayed picture that represents the first referent in their utterance (Bock et al., 2003). Speakers also tend to look at objects immediately before naming them (Griffin and Bock, 2000). In contrast, van der Meulen et al. (2001) found that speakers do not always look at the to-be-named objects: speakers allocated less visual attention to given objects than to new ones, and to objects they would later refer to with a pronoun than with a full noun phrase. This suggests that discourse focus (given vs. new) and referential form (pronoun vs. name) modulate the amount of visual attention allocated to a referent during utterance planning.

Here we test whether planning, as measured by anticipatory eye movements, could be related to both reference form and predictability. This is the first study to examine these issues within a discourse context, so we do not have a priori expectations about how these are related. Current theories might suggest that accessible referents would lead to faster planning (e.g., Bell et al., 2009), but this question has not been addressed explicitly.

We will be using looks to panel 2 (the right-most picture, Figure 1) as the metric of early planning, because looks ahead indicate that the speaker is attending to this event and preparing their own upcoming description turn. In all critical items, this event shows the continuing character (the target referent), as well as an action that they are performing. This motivates our use of the entire picture (as opposed to just the character) as our region of interest here. Additionally, there is enough variability in our visual stimuli that this broad region of interest is the most appropriate.

We will also examine fixations to panel 2 in two different windows of time, which represent early vs. late planning. We speculate that when a speaker is still listening and comprehending the first sentence, they may likely be conceptually planning parts of their own upcoming utterance at this point. When this first sentence is finished there is a period of time that is silent right before the speaker begins speaking, and this may be where more lexical formulation planning occurs. These two periods are based on traditional models of utterance planning (e.g., Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998; Wheeldon, 2013); however, we cannot link them directly to phases of production (message conceptualization vs. lexical formulation).

Importantly, this is the first investigation into the time course of planning and how it relates to both thematic role predictability and reference form choice. To test the planning facilitation hypothesis, we will compare effects in both early and late stages of planning to determine if and where planning facilitation occurs, and how it relates to reference form.

One challenge with studying language planning is that we don't know precisely when planning begins, and thus, the direction of causation is not clear. If speakers plan a pronoun early, this choice might guide their looks during the planning periods. Alternatively, if speakers happen to look more at panel 2, this attention itself might drive the choice of pronoun. Nevertheless, we think it is plausible that events occurring earlier in time are likely to influence events occurring later in time. We therefore take this as our starting point. Our analyses look first at the effect of the discourse context, which occurs earlier than the participants' response. How does the context affect both planning measures and utterance form? Then we look at the effect of planning, which

is logically prior to the reference, and ask whether either early or late planning measures predict variation in reference form.

A summary of our predictors and measures is shown in Figure 6. Using a variation on Rosa & Arnold (2017)'s paradigm, we expect to replicate the effect of thematic roles on reference form. We additionally test three measures of planning: early looks, late looks, and latency. These measures allow us to test three questions:

- Question 1. Does thematic role predict production planning measures (latency, early looks, and late looks)?
- Question 2. Does each of the production planning measures predict reference form choice?
- Question 3: Can the predictability effect on reference form be explained in terms of utterance planning?
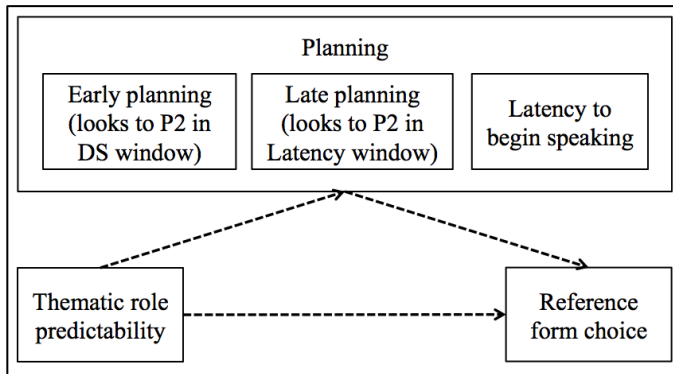

Figure 6. Visual summary of the measures used.

If thematic role affects planning, we expect the goal continuation contexts to elicit shorter latencies (as in previous studies), and perhaps more looks to the target panel during either early or late planning regions. If planning facilitation drives reference form choice, we expect reference form to vary as a function of either latency or anticipatory fixations. Finally, if planning explains predictability effects on reference form, then we would expect planning to have a stronger effect on reference form than thematic role. Planning effects may occur for early measures of planning, late measures of planning, or both.

2   Methods

2.1   Participants
54 native English speakers at the University of North Carolina at Chapel Hill participated for course credit. 13 of these participants used only NP descriptions to refer to the target on critical trials, so these participants were excluded from analyses, leaving 41 participants with reference form variation. 6 of these participants were excluded due to mis-calibration of the eyetracker, and 1 participant was excluded for audio technical failure. This left 34 participants in the final analyses (19 females, and 15 males).

2.2   Materials and Design
A narrative production paradigm designed by Rosa & Arnold (2017) was modified for use with eyetracking for the current study (for stimuli see http://jaapstimuli.web.unc.edu/transfer-verb-stimuli-and-paradigm/). In the primary task, participants viewed pairs of pictures (like Figure 1) and listened to a sentence describing the first (left) picture. Their task was to describe the second

picture, and fixations to the first panel represented attention to the discourse context, while fixations to the second panel represented early planning of the target utterance.

As in Rosa and Arnold's task, participants first watched a background video that described the story context. They were given the role of tabloid photographer, and informed that they witnessed a murder in a mansion and they happened to capture pictures of the events from that day. They learned about six characters: three males (the duke, the butler, the driver) and three females (the duchess, the maid, the chef). Their task was to describe these pictures to a "detective" in order to help solve the crime. One critical change in this study was that the detective's sentences were recorded, instead of spoken live (as in Rosa & Arnold, 2017). The recorded voice described the first of each picture pair, and participants were encouraged to "continue the story" when describing the second of each picture pair. Another critical change was that the instructions emphasized storytelling. In a similar task, Zerkle & Arnold (2016) found that some participants ignored the context and produced only descriptive noun phrases. Here we emphasized storytelling in order to encourage more connected discourse, and thus elicit greater variation in referential forms. The experimenter instructing each participant also gave two example trials (from filler trials of the main experiment), in which they used both connector words and pronouns in each continuation.

The story in the main experiment consisted of 53 "evidence photos": 29 fillers and 24 critical items. In the critical items, the first picture and sentence described a transfer event between two characters (from a source to a goal). The second panel only pictured the target character, which was either the goal or the source of the previous sentence (manipulated between items). The grammatical placement of the target character in the detective's sentence was manipulated between-subjects, such that half of the participants heard, "The Duchess handed a painting to the Duke" and half heard, "The Duke received a painting from the Duchess." Thus there were four possible conditions of character continuation: goal/subject, source/subject, goal/non-subject, source/non-subject. However, the trials from both non-subject conditions were not included in analyses here (see section 3.2 for explanation).

Filler trials pictured between one to three characters. All trials were presented in the same order for all participants to create a coherent story sequence. Character gender was controlled, such that half of the critical trials had two characters of the same gender, and half had two characters of different gender.

2.3   Procedure

Research assistants fitted each participant with an Eyelink II head-mounted eyetracker (SR Research) and a headset headphones/microphone. First participants viewed a slideshow in PowerPoint, which played the background video describing the story and characters, and then they viewed all 53 picture pairs silently in order (5s per pair). The purpose of this initial preview was to familiarize the participant with the series of the events and the outcome of the story. This imitates the characteristics of natural language production, in which speakers usually convey information they already know. After this, they heard two sample trials from the experimenter, and were instructed to continue the story as much as possible when they described each picture. Calibration of the eyetracker occurred right before the main experiment began. Before each trial, participants viewed a drift correct screen and pressed the space bar to continue. Then the pair of pictures appeared on the screen next to each other and remained there for the duration of the trial. After a preview period (average of 665ms), the context detective sentence played over the headphones. Participants spoke their description of the second picture into the microphone, and then used the mouse to click on a green circle in the bottom right corner to move on to the next

trial, which began with another drift correct screen. See Figure 7 for a schematic diagram of the preview and trial procedures.
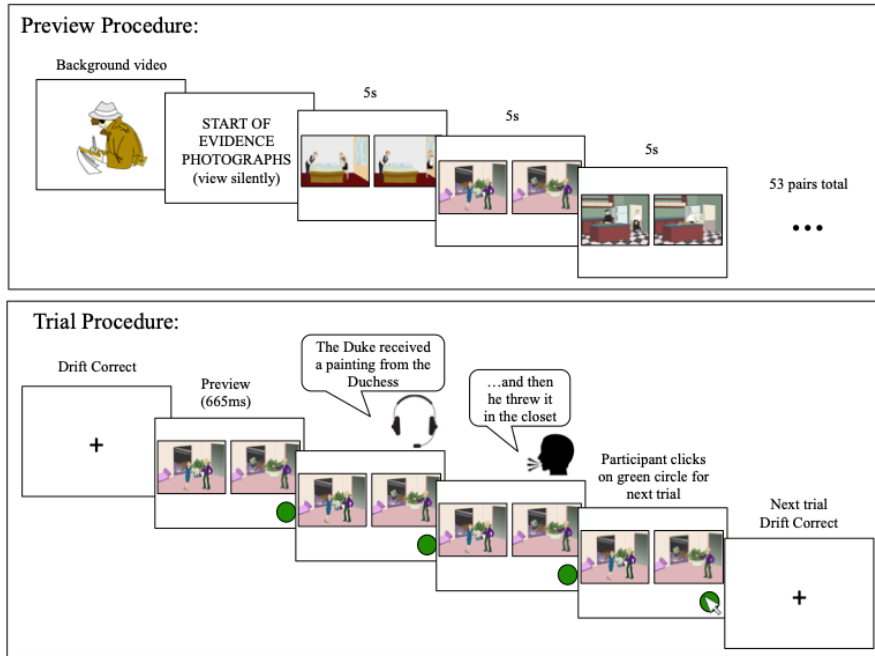

Figure 7. Schematic diagrams of the preview procedure and an example trial procedure.

3   Results and Discussion

In this section, we first describe the general procedure we used to analyze the data from this study. We then report the linguistic, planning, and reference form results.

3.1   General analytic procedure

3.1.1 Response coding

Trials were excluded if the participant did not refer to the target character in the subject position of their utterance. We excluded trials in which the speaker referred to the wrong character in the response (n=9), leaving 373 trials to be audio coded. Research assistants transcribed utterances and coded responses for type of reference (pronoun, zero, NP description) and use of connector before the reference (e.g., *and, then, and then, now*). For the analyses, we only analyzed the subject continuation trials (n=382), excluding the non-subject continuation trials (n=379) (see section 3.2).

3.1.2 Audio data coding

Four research assistants analyzed the audio files in Praat (Boersma & Weenink, 2015). They segmented each trial in order to measure latency to begin speaking, which was calculated as the time in milliseconds between the end of the context sentence and the beginning of fluent speech. Any disfluencies prior to the response were considered as part of planning time, and thus were included in this latency measure (N=10). Outliers were excluded, where outliers were trials where the latency was less than or greater than 3 standard deviations from the grand mean (N=3). We also calculated track loss on a trial level, where trials with greater than 33% of looks in either time window were classified as No Data (blinks, looks not to the computer screen, etc.). There

were 4 trials with < 33% track loss in the DS window, and 5 trials in the Latency window. Thus, 361 trials were included in statistical analyses.

3.1.3 Dependent measures

Our analysis of reference form examined the binary contrast between reduced (pronoun and zero) and unreduced (description) forms. Pronouns and zeros were combined because they play a similar pragmatic role, and are both more common when the referent is accessible or predictable. Zeros in our task represent cases of syntactic coordination, e.g. *The duke received the basket from the duchess…. and Ø threw it down the hallway.* Some studies instead exclude zero continuations from analysis (e.g., Rosa & Arnold, 2017), especially when they are rare in the dataset. Here, however, doing so would exclude around 25% of the data, and misrepresent the rate of using reduced expressions. Moreover, an examination of pronoun vs. zero trials suggests that this decision did not change our findings, in that numerically similar patterns were found for pronouns and zeros. If zeros are excluded from the analyses, similar results obtain.

Our analysis of latency used the log of the latency as a dependent measure. For analyses of gaze, eye position was sampled every 4 ms, but the data were converted to samples at every 20 ms prior to analyses for faster processing (McMurray, 2002). Eye movement data is presented in terms of looks, where a look is defined as a fixation grouped together with the prior saccade. This practice is often used in language processing studies, where listeners are directing their attention to one referent and not another (e.g., Arnold, Hudson-Kam, & Tanenhaus, 2007; Arnold & Lao, 2015; Huettig, Rommers, & Meyer, 2011; McMurray, Tanenhaus, & Aslin, 2009). We used two areas of interest: Panel 1 and Panel 2 (see Figure 1). Ports used for analyses were rectangles exactly around each picture. Any looks to areas besides these were considered "other." Average looks are presented graphically here for ease of interpretation, but empirical logits of looks were used in all statistical models. Empirical logits were calculated for both panels using the following formula based on Barr's (2008) methods:

$$LOG\left(\frac{Sum\ of\ looks\ to\ Panel\ 2\ + 0.5}{Total\ sum\ of\ looks - Sum\ of\ looks\ to\ Panel\ 2\ + 0.5}\right)$$

3.1.4 Model-building procedure

Generalized linear mixed effects models were used to account for dependencies in repeated measures using SAS 9.4. Proc GLIMMIX was used for analyses of dichotomous dependent variables with a logit link, i.e. for all analyses of reference form. Proc MIXED was used for analyses of continuous dependent variables, i.e. for all analyses of eye gaze and response latency. We constructed models of each dependent variable with random intercepts of participant and item to account for nesting. Binary predictors were effects coded, such that each variable as reported in the models below are comparison group vs. reference group. All predictor variables were grand-mean centered. Random slopes of all included variables by participant and by item were included when appropriate, but if a slope was estimated to be zero it was excluded (Searle et al., 1992). The final fixed and random effects for each model are reported below each table.

3.2 Linguistic measures

Speakers rarely used reduced expressions for non-subject continuation trials (10%), so only subject continuation trials were included in all further analyses. Therefore, all subsequent analyses examined only the subject trials. See Figure 8 for a breakdown of pronoun/zero trials by non-subject and subject conditions.

Within the subject continuation trials, the descriptive statistics by thematic role condition are reported in Table 1.

Variable Averages:	Goal condition	Source condition	Overall
Detective sentence length (ms)	2453.73	2242.10	2351.14
Latency to begin speaking (ms)	1454.71	1670.13	1559.14
Pronoun/zero use	64.52%	45.14%	55.12%
Pronoun use	37.10%	22.29%	29.92%
Zero use	27.42%	22.86%	25.21%
Looks to P2 in DS window	47.65%	43.22%	45.50%
Looks to P2 in Latency window	75.88%	68.14%	72.13%

Table 1. Descriptive statistics by goal/source condition, subject continuation trials only. Percentages represent the total number of trials in which an event occurred in each thematic role condition (and overall).

First, we built a model of pronoun/zero use that showed that speakers used significantly more pronouns/zeros for goal continuations, (see Table 2, Figure 8). This finding replicated the effects found in both Rosa & Arnold (2017) and Zerkle, Rosa, & Arnold (2015).

Effect	Estimate	SE	DF	t-value	p-value
Goal vs. Source continuation	1.1739	0.4486	19.88	2.62	0.0166

Table 2. Critical predictor of pronoun/zero use*.
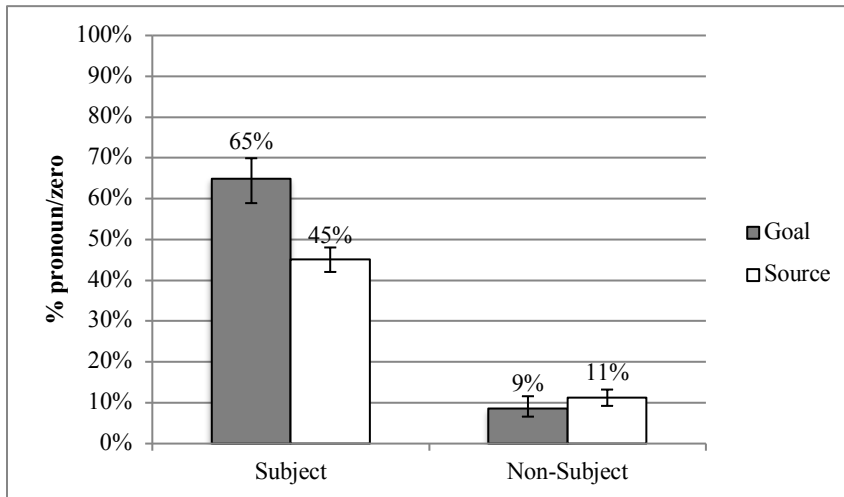* Random intercepts of participant and of item.

Figure 8. Reference form choice rates by grammatical role and thematic role. This includes 373 trials in the subject conditions, and 364 in the non-subject conditions. Error bars are standard error of the subject means by condition. In the first model, the main effect of thematic role is seen in the left bars (only using subject continuation trials).

3.3 Planning measures

Our first question was whether thematic roles would predict variability in each of our three measures of planning: response latency, the speaker's gaze to panel 2 during the context sentence,

and speaker's gaze to panel 2 during the latency period. Looks to panel 2 (which guided the response) were considered evidence of response planning (either message planning or formulation). Following Zerkle & Arnold (2016), we divided each trial into four different time windows (see Figure 9), but our analyses for this study focused primarily on the detective sentence window and latency window, as defined below.

Time windows were calculated starting at the estimated average stimulus image onset. Due to a programming error, there was some variability in the image onset, which was estimated to occur an average of 110 ms after participants responded to the drift correct screen, with a standard deviation of ± 31 ms. The onset of the detective sentence was much less variable, occurring at an average of 665 ms after the estimated picture onset, with a standard deviation of ± 3 ms.

- The Detective Sentence (DS) window: the time between the onset (665ms from stimulus onset) and offset of the context detective sentence.
- The Latency window: the time between the offset of the detective sentence and the onset of the participant's fluent speech.
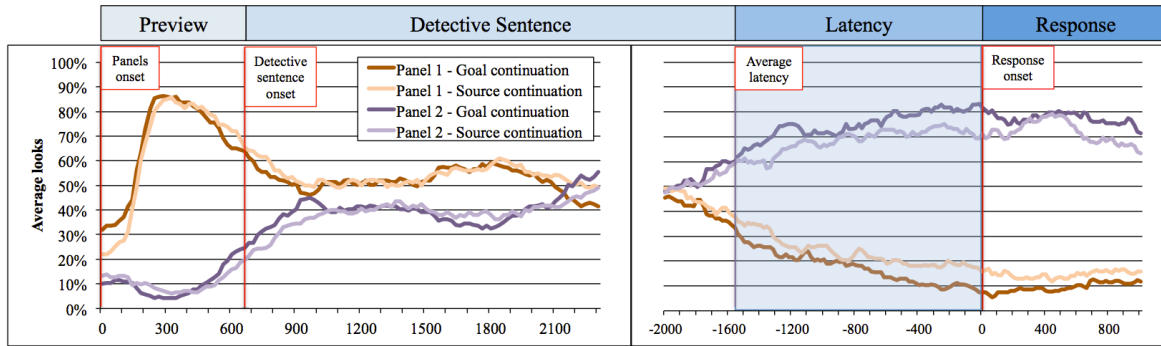

Figure 9. This figure shows the four time windows of a trial, averaged over all subject continuation items. Time is on the x-axis, starting at the beginning of the trial for the left graph, and centered around the onset of participants' response in the right graph. The gray box in the latency window represents the overall average response latency length.

Since we are interested in planning effects, analyses focused on the regions before the response. We have divided these measures into **early planning**: looks to panel 2 during the detective sentence window; and **late planning**: looks to panel 2 during the latency window, and the length of response latency (ms). We test these each in separate mixed effects regression analyses.

3.3.1 Early planning
Speakers were not significantly more likely to look at panel 2 during the detective sentence window for goal continuations (see Table 3, Figure 10). This shows that in answer to Question 1, we did not see strong effects of thematic role predictability on early planning.

Effect	Estimate	SE	DF	t-value	p-value
Goal vs. Source continuation	0.1618	0.107	22	1.51	0.1449

Table 3. Critical predictors of looks to panel 2 in the detective sentence window*.
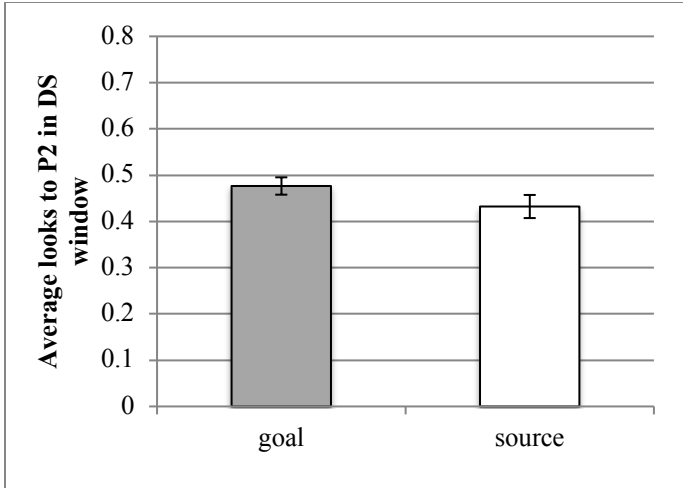* Random intercepts of participant and of item, random slope of goal continuation by subject.

Figure 10. Average looks to panel 2 during the detective sentence region by thematic role condition. Error bars are standard error of the subject means by condition.

3.3.2 Late planning

Response latency was significantly predicted by thematic role, replicating findings from previous experiments using this task (see Table 4, Figure 11; Rosa & Arnold, 2017; Zerkle, Rosa, & Arnold, 2015).

Effect	Estimate	SE	DF	t-value	p-value
Goal vs. Source continuation	-0.06905	0.02803	359	-2.46	0.0142

Table 4. Critical predictors of log latency to begin speaking*.
* Random intercepts of participant and of item, random slope of goal continuation by subject.
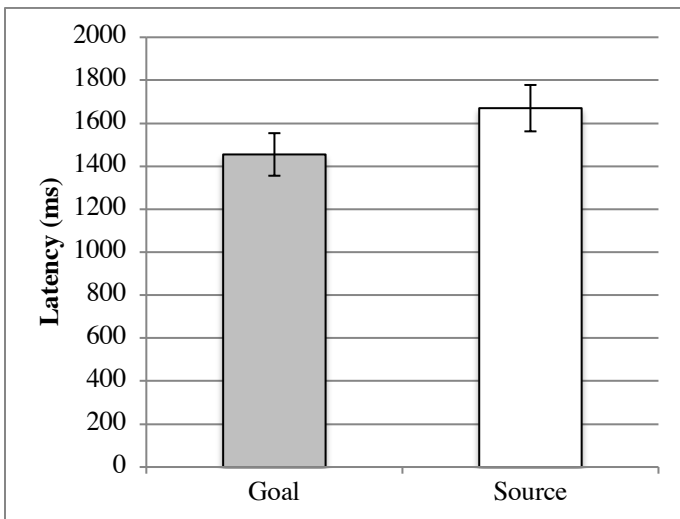

Figure 11. Thematic role effect on response latency. Error bars are standard error of the subject means by condition.

Our next question was whether panel 2 looks during the latency period, which represent a later planning measure, would reflect thematic role condition. We found that indeed, they were significantly influenced by the goal/source manipulation (See Table 5, purple lines in right panel of Figure 9). Speakers looked more at panel 2 in this window for goal continuations than for source continuations.

Effect	Estimate	SE	DF	t-value	p-value
Goal vs. Source continuation	0.2756	0.09201	21.4	3	0.0068

Table 5. Critical predictors of looks to panel 2 in latency window*.
* Random intercepts of participant and of item.

In sum, analyses of our three planning measures show that thematic role influences the duration of the latency period and looks to panel 2 during the latency period, but not looks to panel 2 during the context sentence. Thus, in answer to Question 1, we found that thematic role affected only late measures of planning, and not our early measure of planning. This suggests that predictability has an effect on utterance planning, but only immediately before utterance articulation.

3.4 What explains referential form choice?

Our second question was whether our planning measures would help explain speakers' choices of referential form. To assess this, we examined the three dependent measures from the previous section (early looks, late looks, and latency duration). We first tested the correlations between these three variables (using the transformed values) to assess multicollinearity. Early looks and late looks were marginally positively correlated, ($r=0.10$, $p=0.052$); early looks and latency duration were significantly negatively correlated ($r=-0.23$, $p<.0001$); and late looks and latency duration were significantly negatively correlated ($r=-0.17$, $p=0.001$). All three of these correlations reflect only small effect sizes (Cohen, 1988), given our large sample sizes. This suggests that multicollinearity is not an issue in a model with all three variables.[3]

As shown in Table 6, we found that only early looks significantly predicted reference form choice, such that more looks to panel 2 in the DS window increased the likelihood of using a pronoun/zero (see Table 6).

Effect	Estimate	SE	DF	t-value	p-value
Looks to P2 in DS window	0.5813	0.2666	357	2.18	0.0299
Looks to P2 in Latency window	-0.04723	0.187	41.46	-0.25	0.8018
Log Latency (ms)	-0.8283	0.7756	357	-1.07	0.2863

Table 6. Utterance planning predictors of pronoun/zero use*.
*Random intercepts of participant and of item, random slopes of latency by subject and by item.

3.5. Do planning measures account for the effect of thematic role on reference choices?

Our final question was how these planning measures, together with our thematic role manipulation, account for choices in reference form. Specifically, we wanted to know whether the planning measures would account for the effect of thematic role. We focused on early looks as a planning measure, since this was the only one to predict reference form choice in the previous model. We therefore built a model with reference form (pronoun/zero vs. description) as the

---

[3] Separate models were also run with each individual predictor, and the results from each showed a similar pattern as the full model (only early looks significantly predict reference form).

dependent measure, and with three predictors: thematic role predictability, early looks, and the interaction between the two. Thematic role and early looks were significantly positively correlated, however this significance is contingent upon the large sample size and the effect size is rather small (r=0.124, p=0.019) (Cohen, 1988).

As shown in Table 7, we again found a significant effect of goal continuation (more reduced forms for goals); and also a significant effect of looks to panel 2 in the DS window (more reduced forms for trials with more anticipatory looks). The interaction between thematic role and looks was not significant.

However, it is not the case that the thematic role effect can be entirely explained in terms of early planning. If it were, we would see the thematic role effect disappear in the presence of the DS-window looks predictor. Thus, in answer to our last question, planning does not account for the bias to use more reduced expressions for goals than sources. Instead, we see two independent effects of planning and thematic role.

Effect	Estimate	SE	DF	t-value	p-value
Goal vs. Source continuation	1.173	0.4238	20.09	2.77	0.0118
Looks to P2 in DS window	0.7468	0.2912	357	2.56	0.0107
Goal continuation*Looks to P2 DS	0.7385	0.5503	357	1.34	0.1804

Table 7. Critical predictors of pronoun/zero use, including eye movement predictors*.
* Random intercepts of participant and of item.

3.6. Empirical summary
In sum, we found that goal thematic roles lead to both more reduced forms, and evidence of greater planning during the latency period, as indicated by both shorter latencies and greater anticipatory looks to the target panel. In addition, we found that one of our planning measures (early anticipatory looks) predicted the use of reduced forms. However, these findings are critically independent. The effect of thematic roles on planning occurred only for the **late** planning measures, whereas the effect of planning on reference form only occurred for the **early** planning measure. In addition, including both early looks and thematic role as predictors of reference form revealed that they are independent predictors.

4   General Discussion
In sum, this study revealed three main findings: a) thematic role predictability affects late measures of planning, b) early planning supports the use of reduced forms; and c) planning does not account for predictability effects on reduced forms. We will discuss each in turn.

4.1   Thematic role predictability affects late measures of planning
In answer to Question 1, this study found that thematic role predictability only affected planning during the latency window and the latency measure itself. This is consistent with earlier reports that response latencies are shorter for goal vs. source continuations (Rosa & Arnold, 2017; Zerkle, Rosa, & Arnold, 2015). We found that eye movements during the latency window reflected thematic role predictability, such that speakers looked more at the target stimulus (panel 2) in the goal condition than in the source condition. In other words, only at this late period of planning does the predictability of goal continuations increase looks to the target picture. It may be that speakers are more likely to look back to panel 1 for source continuations in this window because these items require additional comprehension of the context (beyond the length of the detective sentence) before planning.

4.2   Early planning supports the use of reduced forms

In answer to Question 2, we found that only early planning supports the use of reduced referential forms. Importantly, this study is the first to examine reference production within a discourse context. Van der Meulen et al. (2001) also examined eye movements during the production of pronouns vs. noun phrases, but their study didn't test natural language production within a discourse context. We found that early planning affects reference form (both pronouns and zeros) within a natural discourse context. This finding is consistent with Arnold & Nozari (2017) who also found that planning supported pronoun/zero use, in a different paradigm.

This finding is consistent with classical models about reference form, which suggest that the discourse context influences decisions about when to use a pronoun. But critically, this is one of the first studies to link these decisions with inter-trial variation in production planning processes. When speakers are listening to the detective sentence, sometimes they begin planning earlier, as indicated by looks to panel 2. Those are the trials where they tend to use pronouns and zeros. It may be that advance planning is especially helpful when it co-occurs with processing the discourse context (Arnold & Nozari, 2017). Doing so helps establish links between the events, and increases the speaker's desire to communicate these links with their linguistic choices, like the use of pronouns.

4.3   Early planning does not account for predictability effects

In answer to Question 3, we found that early planning does not account for the effects of thematic roles on reference form. This is not consistent with the idea that early planning is the sole driving force. The relationship between predictability, planning, and reference form is complex, but this is the first step towards parsing the functions of early vs. late planning mechanisms for the production of reference form. In sum, planning facilitation is not the reason why predictability affects reference form, so there must be another characteristic of predictability that is responsible for the pattern of behaviors observed here.

4.4   Conclusions

The three main findings from this study rule out a production processing-based account, because planning facilitation does not directly account for pronoun choice. What, then is the likely explanation for predictability effects on reference form? Our speculations on this question build on the frequent observation that referential forms are selected on the basis of the referent's cognitive status, i.e. its discourse salience or accessibility (Ariel, 1990; Chafe, 1994; Givón, 1983; Gundel, Hedberg, & Zacharaski, 1993). This classic view suggests that reduced forms are triggered by the representation of the discourse entity itself (Arnold, 2016). This view is consistent with our findings that early planning supports the use of pronouns and zeros, because we assume that participants are building a discourse representation of each story during the first part of the trial. When that representation is strong, speakers may have greater mental resources to begin pre-planning, and therefore look to panel 2 earlier. Thus, pre-planning does affect reference form, but only pre-planning that occurs during the first part of the trial.

By contrast, the effect of thematic role predictability emerges later. We only see the effect of thematic roles on later measures of planning, and these planning measures are unrelated to reference form. Instead, thematic role must affect reference form through some other mechanism. The thematic role effect points to a property of references that derives from their participation in events, as opposed to questions about how activated or accessible each referent is alone. This raises the possibility that pronouns and zeros are used in cases where there are strong relationships between utterances. That is, goal references are predictable, and as a result are more tightly connected at a conceptual level. In support of this, Rosa & Arnold (2017) asked a group of subjects to rate the pairs of events in their stimuli (also used in the current study). They

found that the two events were perceived as more related in goal continuation than source continuation items. This suggests that speakers may perceive the goal continuations as more connected, which may have downstream effects on reduced form use. Further research is needed to better understand the relationship between thematic role predictability and reference form choice.

Nevertheless, the current findings support the importance of early planning on discourse production. When speakers pre-plan, they are likely to connect events conceptually. In an unrelated finding, thematic roles may also lead to greater conceptual connection, but this effect is not driven by early planning. Both situations lead to the selection of linguistic forms that signal coherence, such as pronouns and zeros.

Acknowledgements

References

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

Ariel, M. (2001). Accessibility theory: An overview. In Sanders, T., Schliperoord, J. and Spooren, W. (Eds.), *Text representation*. Amsterdam: John Benjamins (Human cognitive processing series), pp. 29-87.

Arnold, J. E. (1998). *Reference Form and Discourse Patterns*. Doctoral dissertation, Stanford University, Stanford, California.

Arnold, J. E. (2001). The effect of thematic role on pronoun use and frequency of reference continuation. *Discourse Processes, 31*, 137–162. https://doi.org/10.1207/S15326950DP3102_02

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

Arnold, J. E. (2016). Explicit and emergent mechanisms of information status. *Topics in Cognitive Science*, *8*, 737–760. http://doi.org/10.1111/tops.12220

Arnold, J. E. (2017). Latency analysis for Rosa & Arnold (2017). Technical Report #1. UNC Language Processing Lab, Department of Psychology & Neuroscience, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina. http://arnoldlab.web.unc.edu/files/2017/08/Arnold_TechReport1_2017.pdf

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*(4), 521-536. https://doi.org/10.1016/j.jml.2006.09.007

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say *thee uh* you are describing something hard: the on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(5), 914–930. http://doi.org/10.1037/0278-7393.33.5.914

Arnold, J. E., & Lao, S.-Y. C. (2015). Effects of psychological attention on pronoun comprehension. *Language, Cognition and Neuroscience*, *30*(7), 832–852. http://doi.org/10.1080/23273798.2015.1017511

Arnold, J. E. & Nozari, N. (2017). The effects of utterance planning and stimulation of left prefrontal cortex on the production of referential expressions. *Cognition*, *160*, 127–144. https://doi.org/10.1016/j.cognition.2016.12.008

Arnold, J. E. & Watson, D. G. (2015). Synthesizing meaning and processing approaches to prosody: performance matters. *Language, Cognition, and Neuroscience, 30,* 88-102. https://doi.org/10.1080/01690965.2013.840733

Arnold, J. E., & Zerkle, S. A. (2019). Why do people produce pronouns? Pragmatic selection vs Rational models. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2019.1636103

Barr D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*(4). 457–474.

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*(1), 92–111. https://doi.org/10.1016/j.jml.2008.06.003

Boersma, P. & Weenink, D. (2015). Praat: Doing phonetics by computer [Computer program]. Version 5.4.09, http://www.praat/org/.

Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W.J. M. (2003). Minding the clock. *Journal of Memory and Language*, *48*(4), 653–685. https://doi.org/10.1016/S0749-596X(03)00007-X

Brown-Schmidt, S., & Konopka, A. E. (2015). Processes of incremental message planning during conversation. *Psychonomic Bulletin & Review*, *22*(3), 833–843. http://doi.org/10.3758/s13423-014-0714-2

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: The flow and displacement of conscious experience in speaking and writing.* Chicago, IL: Chicago University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547-619.

Ferreira, F. & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57–84. https://doi.org/10.1006/ jmla.2001.2797

Fowler, C. A. & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489–504. https://doi.org/10.1016/0749-596X(87)90136-7

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998-998.

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. https://doi.org/10.1016/j.jml.2009.09.001

Givón, T. (1983). Topic continuity in discourse: An introduction. In: Givón, T. (Ed.), *Topic continuity in discourse: A quantitative cross-language study*, 1–41. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.3

Griffin Z. M. & Bock K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4). 274–279.

Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational linguistics*, *21*(2), 203-225.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions. *Language, 69*, 274-307.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *NAACL '01: Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies 2001*, 1–8.: https://doi.org/10.3115/1073336.1073357

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. http://doi.org/10.1016/j.actpsy.2010.11.003

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101-144.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229- 254). Amsterdam, Netherlands: John Benjamins Publishing Company.

Kehler, A. & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics, 39*(1–2), 1–37. https://doi.org/10.1515/tl-2013-0001

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

Kravtchenko, E.; Modi, A.; Demberg V.; Titov, I., & Pinkal. M. (2017). Does referent predictability affect rate of pronominalization? Paper presented at the CUNY conference on human sentence processing, March 2017.

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

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(1), 1-38.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. *Advances in neural information processing systems*, *19*, 849.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313-318.

McMurray, B. (2002). Analysis scripts written in Microsoft Access: EyelinkAnal [computer program]. Version 1.5.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, *60*(1), 65–91. http://doi.org/10.1016/j.jml.2008.07.002

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: eye movements during noun phrase production. *Cognition*, *66*(2), B25–B33. http://doi.org/10.1016/S0010-0277(98)00009-2

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

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production on and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org /10.1017/S0140525X12001495

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107-142.

Rohde, H. & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition, and Neuroscience, 29*(8), 912–927. https://doi.org/10.1080/01690965.2013.854918

Rosa, E. & Arnold, J. E. (2017). Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language*, *94*, 43–60. https://doi. org/10.1016/j.jml.2016.07.007

Searle, S., Cassella, G. & McCullouch, C. (1992). *Variance Components*. New York: Jonh Wiley-Sons. https://doi.org/10.1002/9780470316856

Tily, H. J., & Piantadosi, S. T. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. *Proceedings of the Workshop on the Production of Referring Expressions: Bridging the Gap between Computational and Empirical Approaches to Reference*, 1–8.

van der Meulen, F. F., Meyer, A. S., & Levelt, W. J. (2001). Eye movements during the production of nouns and pronouns. *Memory & Cognition*, *29*(3), 512–521. http://doi.org/10.3758/BF03196402

van Rij, J., van Rijn, H., & Hendriks, P. (2010). Cognitive architectures and language acquisition: a case study in pronoun comprehension. *Journal of Child Language*, *37*(3), 731–766. http://doi.org/10.1017/S0305000909990560

Weatherford, K., & Arnold, J. E. (under review). Semantic predictability of implicit causality can affect referential form choice. Ms., University of North Carolina.

Wheeldon, L. R. (2013). Producing spoken sentences: The scope of incremental planning. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech production and perception: Vol. 1. Speech planning and dynamics*. Frankfurt, Germany: Peter Lang.

Zerkle, S. A. & Arnold, J. E. (2016). Discourse attention during utterance planning affects referential form choice. *Linguistics Vanguard, 2*(s1). DOI: https://doi.org/10.1515/lingvan-2016-0067

Zerkle, S. A., Rosa, E. C., & Arnold, J. E. (2015). Do addressee gestures influence the effects of predictability on spoken reference form? Presented as a poster at the CUNY 2015 Conference on Sentence Processing, USC, Los Angeles, CA. March 19-21, 2015.

Zerkle, S. A., Rosa, E. C., & Arnold, J. E. (2017). Thematic role predictability and planning affect word duration. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1): 17, 1-28. DOI: http://doi.org/10.5334/labphon.98