RUNNING HEAD: Planning and stimulation of left PFC affect reference

The effects of utterance timing and stimulation of left prefrontal cortex on the production of referential expressions

Jennifer E. Arnold[a], Nazbanou Nozari[b]

Department of Psychology, University of North Carolina at Chapel Hill

Department of Neurology, Department of Cognitive Science, Johns Hopkins University

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

ABSTRACT

We examined the relationship between the timing of utterance initiation and the choice of referring expressions, e.g., pronouns (*it*), zeros (*...and went down*), or descriptive NPs (*the pink pentagon)*. We examined language production in healthy adults, and used anodal transcranial direct current stimulation (tDCS) to test the involvement of the left prefrontal cortex (PFC) in the timing of utterance production and the selection of reference forms in a discourse context. Twenty-two subjects (11 anodal, 11 sham) described fast-paced actions, e.g. *The gray oval flashes, then it moves right 2 blocks*. We only examined trials in contexts that supported pronoun/zero use. For sham participants, pronouns/zeros increased on trials with longer latencies to initiate the target utterance, and trials where the previous trial was short. We argue that both of these conditions enabled greater message pre-planning and greater discourse connectedness: The strongest predictor of pronoun/zero usage was the presence of a connector word like *and* or *then,* which was also tended to occur on trials with longer latencies. For the anodal participants, the latency effect disappeared. PFC stimulation appeared to enable participants to produce utterances with greater discourse connectedness, even while planning incrementally.

INTRODUCTION

Every time speakers refer, they must choose among various forms of reference. A shape may be referred to by a detailed description (*the pink pentagon),* a pronoun (*it),* or an elliptical (zero) construction, (*…. and went down).* These choices are heavily constrained by the discourse context. Pronouns tend to refer to recently mentioned and accessible entities (Ariel, 1990, 2001; Arnold, 1998, 2008, 2010; Chafe, 1976; Gundel et al., 1993; Givon, 1983), and elliptical (zero) constructions like *… and Ø moves right two blocks* are usually restricted to consecutive utterances with a repeated subject. Yet the context does not provide a categorical, inflexible constraint, and often in the same context multiple forms sound acceptable. This leads to the impression that the difference between reduced and explicit expressions is one of preference, or degree of appropriateness, leaving open many questions about the cognitive mechanisms that drive referential variation. We specifically hypothesize that variability in referential form may be related to variability in the timing of utterance initiation, which reflects the degree to which the message may be pre-planned.

The current study examines this hypothesis in two ways. Behaviorally, we investigate the relationship between the timing of utterance initiation and reference form, testing the hypothesis that reduced expressions occur more often under timing conditions that support discourse connectivity, such as greater message pre-planning. Neurally, we examine how reference form is influenced by the stimulation of the prefrontal cortex (PFC), an area shown to be involved in executive function generally, and utterance planning specifically. We tested these questions in healthy adults, and used anodal transcranial direct current stimulation (tDCS) over the left dorsolateral prefrontal cortex (L-DLPFC). Half the participants performed the task under a sham setup; results from this group established the role of planning processes in language production,

in the absence of stimulation. We then examined the performance of participants under

stimulation to identify ways in which stimulation changes performance.

To our knowledge, our study provides the first test of the relation between the timing of

utterance initiation and reference form production. This is also the first study to examine the role

of PFC on reference production with tDCS.

Why might the timing of utterance initiation affect reference form?

This study examines the time needed between observing an action and describing it

verbally, which is of interest because it reflects the amount of time potentially spent pre-planning

an utterance. Imagine that a speaker observes a blue square moving and says *The blue square*

loops around the pink triangle. Production of this utterance requires several steps, including a)

identifying the message to be communicated, i.e. the shape that is moving, the action it is

performing (looping), and the shape it is looping around, b) selecting the words for each phrase,

c) building a syntactic structure, d) building a phonological representation, and e) generating a

phonetic representation (e.g., Levelt, 1989). For our purposes, the critical step is message

planning.

Speakers tend to plan each element of an utterance in sequence, but there is variability in

how much of the message is planned before articulation begins (Konopka, 2012). Speakers may

begin their utterance as soon as they identify the referent of the subject NP, for example saying

The blue square while they figure out the action and plan the rest of the message in parallel with

speaking. This approach would reflect a highly **incremental** mode of message planning and

speaking. Alternatively, speakers may **pre-plan** a larger segment of the message, where they

wait until a chunk of the message (or even the entire sentence) is planned before initiating the

utterance. These two alternatives represent extremes on a continuum. Pre-planning the message

does not require that linguistic formulation is also pre-planned, but message planning is at the very least a necessary condition for linguistic formulation to begin. In addition, there is known variation in the scope of verbal pre-planning (Ferreira and Swets, 2002; Griffin, 2003; Konopka, 2012; Meyer, Belke, Haecker, & Mortensen, 2007; Wheeldon & Lahiri, 1997; Schriefers & Teruel, 1999; Wagner et al., 2010).

The degree of message pre-planning is influenced by two competing pressures on speech production. On one hand, the social demands of language production induce an implicit goal of fluency (Clark & Fox Tree, 2002, Clark & Wasow, 1998). Fluency requires the speaker to plan enough of a phrase ahead of time in order to utter it without pausing. On the other hand, the social demands of language also limit the time the speaker can take to plan, in that long delays can signal that the speaker is finished, or can be perceived as nonfluent.

The intuition behind our study is that the timecourse of message planning has implications for the conceptual links between utterances in a connected discourse. When speakers pre-plan one message while articulating the previous sentence, the parallel processing may encourage conceptual ties between utterances. The reason for this conceptual overlap stems from the staged nature of language production. A message is planned conceptually, and then encoded linguistically, before it is sent to the articulation stage. While message planning and linguistic encoding may overlap, the articulatory buffer has a limited capacity of 1-2 words (Levelt, 1989; Garrett, 1975). This means each component of the message must be kept active until shortly before it is articulated. In a task like this one, where the actions determine message planning, and where the actions follow each other within seconds, the speaker may often be holding one sentence in memory while viewing the movement for the next sentence and planning

it conceptually. If planning of the second sentence happens while the first sentence's message is still active, the two messages have to be active simultaneously and are more likely to be linked.

Our first hypothesis is that this conceptual integration facilitates normal processes of representing discourse relations, thus increasing the speaker's tendency to produce linguistic indicators of discourse connectedness, such as pronouns. We compared this hypothesis with two alternate possibilities. One possibility is that pre-planning the target utterance does not in fact support the use of pronouns and zeros, because it requires dual-tasking, i.e. speaking one utterance while planning the next one. This may be difficult and lead to interference between articulation and planning processes. Such interference may cause difficulty remembering the discourse context, inhibiting the production of reduced expressions, contrary to what was predicted in the first hypothesis. Another possibility is that that pre-planning does indeed increase the proportion of reduced forms, but not because of conceptual integration. Rather, pre-planning may instead avoid disfluency, which tends to suppress the production of reduced expressions.

Next we describe our task and measures, before we show how our hypotheses make predictions within this task.

The Moving Objects paradigm and predictor variables

Participants described visual events in the Moving Objects paradigm (Nozari, Arnold, & Thompson-Schill, 2014; Figure 1; for a video example, see https://arnoldlab.web.unc.edu/publications/supporting-materials/arnold-nozari-2017/Our analyses focused on the contrast between modified noun phrases like *the pink pentagon*, compared with reduced expressions (pronouns and zeros). Notably, all trials in our analysis

occurred in a discourse context that supported the use of pronouns and zeros, i.e., trials on which

the same shape moved on the previous trial. Thus our focus is not on the contribution of the

discourse context per se, but rather on how form choice varies as a function of the timing of

speech with respect to the timing of the stimulus actions.

Example sequence:

The yellow pentagon moves down 1 block
<u>The yellow pentagon</u> flashes
Then <u>it</u> jumps over the yellow square
The pink pentagon loops around the yellow square
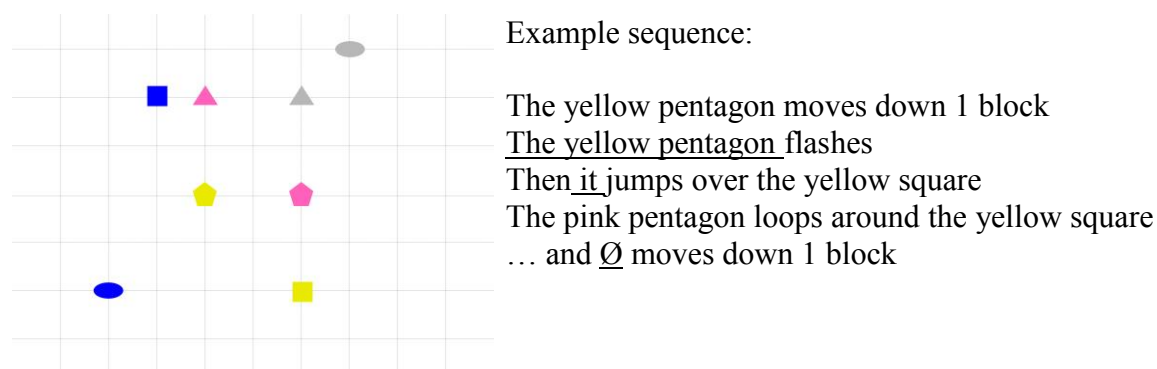… and <u>Ø</u> moves down 1 block

Figure 1. Example display and sequence of action descriptions for one set. Utterances involved

in the critical analysis are underlined. The target referring expression is the moving object (i.e.

the subject NP).

As soon as the movement was identified, the participant could begin preparing the

grammatical subject. However, the identification of the action took longer. The action durations

spanned from 1200 to 4820 ms (with 200 ms between each action), leading to variation in the

availability of the verb. All actions except the flash action also contained information that

followed the verb (e.g. how many blocks an object moved, or which object was being jumped

over), and this information also varied in how quickly it was available. Participants were

encouraged to speak as normally as possible, yet the fast-paced nature of the task meant that

speakers could not afford to wait, or else they would likely fall behind on subsequent trials. This

typically meant that speakers began speaking before an action was finished, and were still

speaking when the next action began (see Figure 2 for an illustration).

Utterance 1
latency

Utterance 2
latency

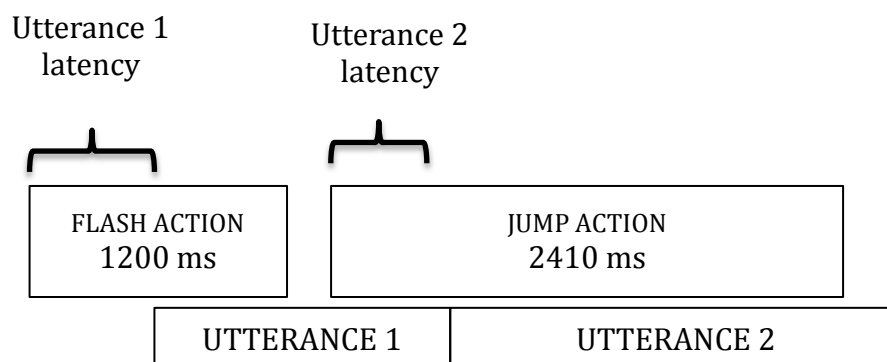| FLASH ACTION 1200 ms	JUMP ACTION 2410 ms
UTTERANCE 1	UTTERANCE 2

Figure 2. Diagram of the timing of actions and utterances in our task. The timing of the

actions was fixed.

The advantage of this task is that it creates a discourse corpus in which we control the

content of each message (which corresponds to an action event), but participants are free to use

wording of their choice. This provides a rich set of linguistic and timing variables, which can be

analyzed in order to understand the role of planning in reference production. Our analysis

examines two linguistic variables: connector use and disfluency, and four timing variables: 1)

latency of target utterance; 2) action duration of target utterance; 3) latency of previous

utterance; 4) action duration of previous utterance. We also examine the relation between the

timing of the previous utterance and the target action onset, to test whether linguistic choices are

different on trials where the action overlaps with the previous utterance. The definition and

predictions of each variable are explained below.

Critical for our investigation, the onset of the action was the earliest point at which

speakers could begin the process of planning any part of the message. Therefore, our first

independent variable measured the latency between action onset and utterance onset, as a

measure of the degree of message pre-planning on the current, target trial (IV-1 = current-latency). The latency period potentially included two component measures: 1) "overlap", i.e. overlap with the previous utterance (which occurred on 92% of the trials), and 2) "planning silence", i.e. the portion of the silence between utterances that occurred after the action onset (see Figure 3 below). These component measures were correlated with the latency (overlap: R= 0.86, p <.0001; planning silence: R = 0.11, p=.005). This demonstrates that latencies tend to be longer when the speaker needs to complete the previous sentence (overlap), and when speakers pause between utterances (planning silence). We used latency as a predictor because it provided the theoretically most comprehensive measure of total possible planning time. The silence measure did not predict reference form significantly by itself (see section on testing alternate effects).

Thus, in our task short latencies reflected trials where the speaker had finished speaking the previous sentence, but had not yet planned the message for the entire target sentence. For example, on one trial the speaker says "The blue square moves down 3 blocks", and then observes an action where the blue square moves left one block. At the start of the second action (moving left), the speaker is still saying the last word of the previous sentence "blocks", and then begins describing the moving-left-1-block action only 671 ms after it starts. Thus, this trial with a short latency corresponds to relatively little overlap with the previous trial. Given that the moving-left-1-block action takes 1210 ms to complete, this also means that speaker had not yet seen the number of squares the object was moving on the target trial. By contrast, on trials with long latencies, subjects were typically finishing the previous utterance during much of the action movement. While this required subjects to plan the target message while speaking the previous one, it also enabled them to pre-plan. Note that this measure does not indicate which pieces of the message were planned concurrently, nor the scope of grammatical/lexical planning. Instead,

it indicates the amount of information that could have been planned prior to utterance initiation, while finishing the previous sentence.

In addition to current-latency, we also looked at the duration of the current trial (IV-2 = current-action-duration) as a potential predictor, because longer trials entailed additional information unfolding towards the end of the event, which could potentially change planning strategies. For example, the jump trials (2410 ms) and the loop trials (4820 ms) were initially identical to each other, which means that the action was ambiguous. In the moves-3-blocks items (3610 ms), it wasn't clear how many blocks the object moved (1, 2 or 3) until the end of the trial.

Finally, we investigated two indices of performance related to the previous trial. As shown in Figure 2, current-latency typically included the time spent finishing the previous utterance, which could potentially be affected by two measures: how long the previous action was (IV-3 = previous-action-duration), and how long speakers took to begin speaking on the previous trial (IV-4 = previous-latency). It is thus important to account for the contribution of these variables to current-latency in the analyses. However, it is possible that previous-trial measures also have an influence on reference-form production independently of their effect on current-latency. For example, when the previous action is long, message planning is usually finished long before the event is over. For example, while "looping" takes 4800 ms, all the relevant information for message planning is available by about 2400 ms. This introduces a conceptual gap between the two events, not because of the planning strategy the speaker chose to use, but because of the inherent timing of the events.

In sum, this task allowed us to examine the relationship between the timing of both stimulus actions and utterance production, on the one hand, and variation in linguistic form, on

the other. Below, we discuss how these variables can be used to test three hypotheses regarding

reduced form generation.

Hypotheses and predictions

The discourse connectivity hypothesis.

The discourse connectivity hypothesis predicts that variation in the timing of message

planning affects reference form by increasing the speaker's ability or inclination to link an

utterance to the prior discourse context. We first consider the predictions this makes for the

timing variables (Predictions 1-4), and a linguistic variable, connector use (Prediction 5).

Prediction 1: Current-latency (IV-1). As explained earlier, longer current-latencies in this

task mark time windows during which the speakers are concurrently formulating the current

message while finishing the last utterance, whose conceptual message is still kept active. If this

parallel activation of messages helps discourse connectivity, and discourse connectivity is in turn

critical for the choice of referential forms, we would expect longer current-latencies to lead to

more reduced forms. In our task, longer current-latencies also reflect greater pre-planning of the

message, which requires the speaker to conceptualize a larger chunk of information before

beginning to speak. This provides the time for the speaker to consider how the information fits

with the prior context, including anaphoric links, and thus should support discourse connectivity.

By contrast, rapid utterance initiation is likely to lead to more disconnected discourse.

Our predictions about latency rest on the assumption that longer latencies in a task like

this may promote discourse connectivity. This assumption can also be tested by examining

whether speakers produce other markers of connectivity, such as connector words, on trials with

longer latency. Nevertheless, the literature makes few direct predictions about our task, because

few studies have examined pronoun production in connected discourse. Some insight comes

from studies that examine the role of linguistic planning, as opposed to message planning. For example, van der Meulen, Meyer, & Levelt (2001) found that pronouns take less time to plan than descriptions. However, our task differs from van der Meulen's in that the stimulus actions take time to unfold. We predict that in this context, message planning effects will take priority.

Note that the latency measure is under the participant's control. Although these choices may be constrained (e.g., if the speaker is unable to utter a sentence quickly enough), speakers can adopt either an incremental or pre-planning strategy. That is, they choose whether to rush through each utterance, in order to time-lock their descriptions with the event as it unfolds, or pre-plan each message while continuing to articulate the previous utterance.

Prediction 2: Current-action-duration (IV-2). If participants wait to begin speaking until the entire message has been planned, the current action duration may mediate the effect of latency on reference form. However, message planning is also driven by other properties of the discourse context, such as whether the speaker is still uttering the previous sentence. Critically for pronoun/zero production, participants can begin planning the subject portion of their message (e.g., The yellow pentagon…) as soon as the action begins, and can also probabilistically identify the action itself very quickly. For these reasons, current action duration may not have a strong effect on reference form.

Prediction 3: Previous-action-duration (IV-3). Some actions in our task took a long time to unfold, but the speaker had enough information to plan most or all of the utterance well before the action completed. For example, on the loop trials (n=12 per subject), where participants knew that it was a loop (and not a jump) as soon as the object had moved more than halfway around the other object, leaving about 2 seconds more of the action in most cases. On the moves-3 trials, speakers could also plan the entire message as soon as the object moved beyond the second

square. When this happened, the planning of the long event was temporally separated from the following event. That is, the bulk of planning the long action happened at the beginning of the action, but then the speaker had to wait to see the next action, thus imposing a conceptual dissociation between the events. It has been demonstrated that speakers use explicit referring expressions (repeated names or descriptions) to mark a break in the discourse structure, for example when a narrative involves a temporal or spatial shift (McCoy & Strube, 1999; Vonk, Hustinks, & Simons, 1992). If participants had to wait for the action, they may have perceived a discourse break and marked it linguistically with an explicit referring expression. If long events allow speakers to plan the message before the action has completed, it would predict the use of over-explicit referential expressions in trials that come after long events, such as loop (4820) events. Note that this prediction differs from Prediction 1 (current-latency), in that the previous action durations affect the event structure itself. That is, if speakers perceive a break, the appropriate way to present this information is to signal a discourse break linguistically.

If the above prediction is correct, long events may provide enough time for speakers to finish their utterance before the following event. If so, there may be a negative correlation between previous-action-durations (IV-3) and current-latency (IV-1). Including previous-action-durations allows us to test whether any effect of latency is driven by previous action durations.

Prediction 4: Previous-latency (IV-4). Latency to start speaking on the previous trial should not affect reference production unless it also affects the latency of the current trial (IV-1), for example if it causes the previous utterance to continue into the current trial. Therefore, inclusion of this variable in the analyses tests whether the effect of current-latency is imposed by what happens on the previous trial or not.

Prediction 5: Discourse connectors. In predictions 1 and 3, we claimed that both longer

current-latency and shorter previous-action-durations should predict higher rates of reduced

forms, because they promote conceptual ties between events. If conceptual coherence affects

multiple linguistic indicators of coherence, we should be able to see clear links between these

timing measures and linguistic markers of discourse connectivity, such as "and" and "then".

Evidence suggests that speakers are more likely to use pronouns or zeros when they mark

discourse connectivity with words like *and* or *then*, explicitly signaling the discourse

contribution of an utterance (Arnold & Griffin, 2007). If our timing variables really affect

reference form by affecting conceptual coherence, we would expect that they would also reliably

predict the use of linguistic markers of discourse connectivity.

The Interference Hypothesis

The interference hypothesis stands in direct contrast to the discourse connectivity

hypothesis with respect to its predictions about reference form. It posits that parallel articulation

of the previous utterance and the message planning of the current one should lead to interference.

If such dual-tasking imposed a memory load, it might impair other linguistic processes, such as

the production of discourse-appropriate forms. That is, interference may disrupt the conceptual

linking of the current message (i.e., event) with the previous one. This puts the locus of

interference effects on reference form at the conceptual level, not the lexical level (where

pronouns might be considered easier to produce).

This hypothesis makes the opposite prediction to our primary hypothesis about discourse

connectivity, i.e., it predicts that any variables that cause a potential overlap between the

previous and the current utterance should *decrease* the use of reduced forms. In our task, long

latencies tend to occur when the previous utterance had not finished when the target action began,

leading to overlap. Thus, this hypothesis predicts that longer latencies should **decrease** use of reduced forms (in contrast with the discourse connectivity hypothesis). Likewise, if short previous-action-durations or long current-action-durations increase separation between utterances, they may also cause an increase in reduced forms. More directly, this hypothesis can be tested by asking whether reduced forms were more likely on trials with longer "planning silences", defined as the period of time after both the action began and the previous utterance ended, until utterance onset. If interference determines the use of referential forms, longer planning silences should correlate with higher proportions of reduced forms.

The disfluency hypothesis

Unlike the interference hypothesis, the disfluency hypothesis shares certain empirical predictions with the discourse connectivity hypothesis. That is, it also predicts that longer current-latencies should support the use of reduced expressions, but for a different reason, namely by reducing disfluency. In earlier work, we found that disfluent utterances had fewer reduced forms (Arnold et al., 2009). Disfluency is expected to occur more often in sentences with short latencies, which reflect incremental planning (Clark & Wasow, 1998). Clark and Wasow's commit and restore model suggests that speakers initiate utterances before the entire utterance is planned, and use disfluent repetition to restart the utterance. If the commit and restore strategy applies to situations where the message cannot be completely planned, we would predict greater disfluency in utterances with short latencies, and fewer reduced forms in the disfluent utterances. If it does, it could directly influence the production of pronouns and zeros. Thus, we ask whether disfluency mediates any effect of latency and reference form. This hypothesis would be supported if the production of reduced forms is better accounted for by the fluency of the trial than longer latencies.

Testing timing effects on reference production

      At a behavioral level, our goal was to examine how reference form is influenced by the timecourse of message planning and utterance initiation, guided by the hypotheses above. We tested the above hypotheses with data from our sham group, who did not receive stimulation. Our elicitation experiment varied the timing of the actions, while variation in latency to initiate the utterance was under control of the speaker. A summary of the predictors and results for the entire dataset is presented in the Appendix.

The role of the prefrontal cortex in planning and reference form production

      Our second goal was to test how reference form was affected by stimulation of the left PFC. This endeavor represents the first examination of stimulation effects on reference production. While reference production has been extensively studied, most work focuses on the role of the discourse or social context. Our study takes a complementary approach, examining cognitive and processing factors instead. It is known that tDCS stimulation can change the allocation of cognitive resources (Nozari & Thompson-Schill, 2013). Thus if tDCS induces a change in the production of reduced forms or the factors that affect them, this finding will contribute to our goal of linking discourse processing to measures of cognitive and neural processing.

      Our first question was whether stimulation of PFC would affect reference form choices at all. We suspected it would, based on evidence that PFC is involved in allocating resources necessary to plan a sentence from a conceptual event. First, the role of PFC in action planning in humans has long been documented (e.g., Fuster, 1989; Jurado & Rosselli, 2007; Shalice, 1982; Stuss & Benson, 1986). Although PFC is less discussed in the context of sentence planning per se, neuropsychological data suggest that PFC damage affects utterance planning. The typical

profile of patients with Broca's aphasia (who frequently suffer from lesions to the left PFC) is

agrammatical speech that lacks the structure and fluency of well-planned sentences (Damasio,

1992; Dronkers, Plaisant, Iba-Zizen, & Cabanis, 2007). Moreover, an aphasia syndrome

originally described by Luria (1970, 1973) called "dynamic aphasia", now linked to damage to

the left PFC, manifests as markedly decreased spontaneous propositional speech, in spite of

intact comprehension, repetition and picture naming abilities (Robinson, Blair, & Cipolotti,

1998). Thus, there is good reason to expect the involvement of left lateral PFC in sentence

planning, and for stimulation of this region to affect linguistic behaviors that are potentially

affected by planning, such as reference form choice.

Moreover, there is recent evidence that anodal tDCS also affects sentence-level language

production. In a companion study, Nozari, Arnold, and Thompson-Schill (2014) analyzed the

productions in the same dataset as we are analyzing here[1] and found that anodal stimulation of L-

DLPFC reduced the rate of speech errors. Our analysis further supported the idea that anodal

stimulation of left PFC affects planning: Anodal stimulation affected the incidence of speech

errors in a situation where the action was temporarily visually ambiguous. Specifically, the

"jump" and "loop" trials in our task look visually similar up to a point (at which the jump action

stops, but the loop action continues), thus making it likely for speakers to use "loop" instead of

"jump" and vice versa. Participants who received anodal stimulation were less likely to make

this sort of error than those with sham stimulation, indicating superior ability in managing the

ambiguity. One possibility is that they paused in the middle of the utterance, waiting until the

disambiguation point to plan the verb. Another possibility is that participants pre-planned both

the word "jump" and "loop", but anodal participants were better at suppressing the inappropriate

---

1 This analysis had different goals and thus a different set of criteria for inclusion of trials in the
analysis, see Nozari et al. (2014) for details.

one when the action was disambiguated. Both explanations suggest that the stimulation of L-DLPFC can affect message and utterance planning.

We also expected stimulation of PFC to increase the speaker's ability to produce coherent, pragmatically appropriate discourses. As we have argued, our task encourages a certain degree of incremental planning in order to keep up with the pace, and this mode of planning often leads to disconnected and potentially disfluent speech. By contrast, pre-planning promotes discourse connectivity. It is likely that the production of coherent, connected discourses relies on working memory resources, both conceptually and linguistically. Conceptually, the current and previous actions must be represented, as well as the relation between the two. Working memory resources are known to be involved in the calculation of long-distance dependencies (Hartsuiker & Barkhuysen, 2006), which are similar to referential dependencies. It is also likely that message pre-planning may co-occur with pre-planning of the linguistic elements, which also depends on working-memory resources (Martin & Freedman, 2001; Martin & He, 2004). If stimulation of DLPFC enhances working memory, as suggested in past studies (e.g., Fregni et al., 2005; Nozari & Thompson-Schill, 2013), two outcomes are possible: (1) speakers may engage in greater message pre-planning, and thus risk falling behind, or (2) more effectively, they might stick to incremental planning, but use the increased working memory capacity to maintain the connection with the previous utterance even in this mode of planning.

A final possibility is that stimulation of PFC may affect discourse coherence by increasing fluency. Anodal stimulation of the left PFC increases performance of aphasic patients in verbal fluency tasks (which measure the number of words retrieved) and picture naming tasks (Baker, Rorden, & Fridrikkson, 2010; Fertonani et al., 2010; Hamilton, Chrysikou, & Coslett, 2011; Flöel, 2012; Holland & Crinion, 2012; Schlaug, Marchina, & Wan, 2011), as well as the

performance of healthy participants (Iyer et al., 2005). If these word-level effects extend to

utterance-level fluency, tDCS may also facilitate reference production, given evidence that

reduced forms are more frequent in fluent speech (Arnold & Griffin, 2007; Arnold et al., 2009).

In summary, the primary question we asked was whether stimulation of PFC would affect

reference form. If it does, we are further interested in the nature of this effect, and its relationship

to utterance timing measures. One possibility was that stimulation would alter the timing of

utterance production itself, which may have an indirect effect on reference form choice. Another

possibility was that stimulation would not affect the timecourse of planning itself, but would

facilitate the management of the multiple sub-tasks related to utterance production. If so, we

might expect participants to show an increased sensitivity to the discourse context under

stimulation, which would be expected to affect both reference form and other indicators of

discourse connectivity, like the use of connecting words. A final possibility is that it would not

affect discourse connectivity per se, but may affect fluency of production.

To stimulate PFC, we used tDCS; a safe method in which a weak electrical current is

applied to the scalp (Iyer et al. 2005; Kessler et al., 2012). tDCS affects neural threshold of

activation by changing the membrane's resting potential in either an excitatory fashion (anodal

stimulation), which generally facilitates performance, or an inhibitory one (see Nozari, Woodard,

& Thompson-Schill, 2014, for a discussion of facilitatory and inhibitory effects of cathodal

stimulations in cognitive tasks).

To summarize, we used a naturalistic language production task to examine two questions.

We first examined how the timecourse of utterance production was related to reference form

choices, and how both of these variables related to the production of connector words and

disfluency. Second, we asked how PFC stimulation modulated the effects of utterance timing on reference form production, and how it related to both connector production and disfluency.

METHODS

Participants.

Twenty-four participants (13 female) participated in exchange for $20. All were right-handed and native speakers of English, between the ages of 19 and 30. Two participants were excluded from this analysis, one who produced no variation in reference form throughout the experiment (using only definite, modified NPs), and one who asked explicitly about the acceptability of zeros as a reference form during the instruction period.

Materials and Design.

Our referential communication task was designed to elicit both explicit and reduced referential expressions, and encourage incremental planning and production. The actions varied on several dimensions (e.g., discourse context, number of shapes, verb, duration of action), with the goal of creating a variable and challenging task that would both elicit speech errors (Nozari et al., 2014) and create a variety of discourse situations for the current analysis. This yielded a corpus of responses from a highly controlled task.

Participants viewed a series of 134 actions on a Microsoft Powerpoint slideshow, in which eight shapes performed one of five actions, in sequence. There were two versions of the task, which were identical except for the shapes and colors used. Version 1 included a blue square, a yellow square, a blue oval, a grey oval, a pink triangle, a grey triangle, a pink pentagon, and a yellow pentagon (Figure 1). Version 2 included a red rectangle, a purple rectangle, a green trapezoid, a purple trapezoid, a green circle, a brown circle, a brown diamond and a red diamond.

These shapes were situated on a black grid with a white background, as shown in in Figure 1.

Object and color names in the two versions were matched in frequency and number of syllables.

The shapes performed five different actions: flash, wiggle, move (1, 2, or 3 blocks), loop

around another shape, or jump over another shape, although the wiggle action did not occur in

the discourse context analyzed in this paper. Here and throughout, the term N1 refers to the

shape that occurs in grammatical subject position and N2 refers to the non-subject shape, e.g.

The blue oval (N1) *jumps over the blue square* (N2). The trials were divided into 27 sets, with

four to six actions per set. The onset of the first movement in each set was signaled with a beep

that played 30 msec after the action onset. The sequence of actions was variable, and neither the

shape nor the action was predictable based on the previous action. There was a 200 msec delay

between consecutive actions.

Table 1. The timing of each action in the reference analysis. There was a 200 msec delay
between consecutive actions.[2]

Action	Example and action description	Timing from start to completion (msec)	# trials in the given/parallel condition
Flashes	*The gray oval flashes* Object appears and disappears rapidly	1200	2
Moves 1	*The pink pentagon moves down 1 block* Object moves 1 space up, down, left, or right	1210	6
Moves 2	*The yellow square moves left 2 blocks* Object moves 2 spaces up, down, left, or right	2410	4
Moves 3	*The brown circle moves right 3 blocks* Object moves 3 spaces up, down, left, or right	3610	2
Jumps	*The purple trapezoid jumps over the green circle* Moves in a semicircle around another object	2410	10
Loops	*The yellow pentagon loops around the gray*	4820	12

[2] Due to a programming error, one item in this analysis did not have a preceding delay for 12 of
our 22 participants.

	triangle Moves in a full circle around another object		

The analyses in this paper focus on variation in reference to N1, the moving shape, in the 36 trials on which N1 was identical to the moving shape on the previous trial. From a discourse status perspective, this means the referent was given, and mentioned in a parallel syntactic position as its antecedent (i.e., both were subjects). On 4/10 of the jump events and 6/12 of the loop events, the location shape was also repeated from the previous trial. The timing and distribution of these actions are shown in Table 1.

Procedure.

Participants were randomly assigned to a stimulation condition (anodal vs. sham), and were rotated through the stimulus lists (Version 1 vs. Version 2) such that the versions were balanced within condition. Participants sat in front of a 19-inch computer monitor, at a distance of 25 inches.

Participants first completed a 2-minute speed test, which ensured that they were able to keep up with the speed of the task (see Nozari et al., 2014, Appendix A). They then viewed an orientation demonstration, in which they saw each of the eight objects and named them. Following this, they learned each of the five target actions and practiced examples of them. They then watched an example set of four actions and heard a description of them. In the example slide with narration, two NPs and two reduced forms (one pronoun and one zero) were used. When reduced forms were not used, it was inappropriate to do so (e.g., the object has not been mentioned before). Participants then watched the same actions while providing their own descriptions. The experimenter corrected any problems, and then they completed two more

practice sets. If participants did not produce reduced forms during practice, the experimenter

instructed them to "speak naturally, as you would in everyday life", but explicit instruction to use

reduced forms was not provided. As soon as the orientation was over, the main experiment began.

The experiment, including orientation and practice, lasted an average of 17 minutes, and all

participants in the anodal condition completed it during stimulation.

The participant's speech was digitally recorded. Participants were instructed to describe

the events as they were taking place, as a sports reporter would. They were discouraged from

recalling the actions from memory. Each set included 4-6 actions, which proceeded at a fixed

rate and could not be paused. Pilot testing revealed that participants could keep up with this pace,

but it was fast enough to put them under pressure, and elicit speech errors (see Nozari, Arnold, &

Thompson-Schill, 2014). Between sets, participants could take breaks. They pressed a space bar

to begin each of the 27 sets of actions when they were ready.

Direct Current Stimulation

Direct current was generated with a continuous current stimulator (Magstim Eldith 1

Channel DC Stimulator Plus, Magstim Company Ltd., Whitland, Wales). A 1.5 mA direct

current was delivered for 20 minutes via saline-soaked sponge electrodes with a surface area of

25 cm^2, with a 30-second ramp up and ramp down. In the sham condition, stimulation was only

applied for 30 seconds. The location of anode was F3, according to the 10-20 EEG system for

electrode placement. The cathode was located over the right supraorbital region. In a

questionnaire following the experiment, all participants reported that they thought they had been

in the stimulation condition.

Transcription and Coding

The participant's description of the moving shapes was transcribed word for word, including any disfluencies or repetitions, and errors in accuracy or corrections. A different set of research assistants checked the transcription and ensured that all disfluencies were accurately transcribed, including pauses. Pauses were coded impressionistically, and included any break in the intonational phrase, including those that were marked by word lengthening rather than actual silence.

Each trial in the corpus was coded for the following information:

1. Referential form. The shape descriptions were coded as a) description (e.g., *the yellow pentagon*), b) bare description (*yellow pentagon,* with no determiner*)*; c) pronoun *(it)*, d) zero (e.g., *and moved up one block).* The analysis examines the rate of using reduced expressions (pronouns or zeros), since both are used in similar discourse contexts.

2. Disfluency and corrections. Trials were coded as disfluent (containing one or more disfluent element) or fluent. Disfluent elements included filled pauses (*um, uh)*, disfluent pasues, the pronunciation of *the* as "thiy" (rhyming with tree), the determiner *a* as "ay" (rhyming with hay), repeated words or fragments, prolongations, or corrections of speech errors (i.e., inaccuracies with respect to any part of the action). For a full description of the speech error analysis, see Nozari et al. (2014).

3. Action and Timing variables. The current-action-duration (IV-2) and previous-action-duration (IV-3) were variables inherent in the design and did not need to be measured. Current-latency

(IV-1) and previous-latency (IV-4) were identified in *Praat* (Boersma & Weenink, 2013) by locating the beep in the soundfile. Using this as the starting point, we calculated the onset of each action in that block. We also used *Praat* textgrids to identify utterance onsets and offsets, using the waveform and listening to identify the onset of the first word in the utterance. Note that the first word in an utterance was not always the target referential expression, as it sometimes was a disfluency or the word *and*, *then*, etc.

To calculate the current-latency (IV-1), we took the onset of the action as the earliest point when message planning could take place, and report the latency to the onset of the first word of the utterance, whether it is the determiner, pronoun, or verb onset. The top panel of Figure 3 illustrates the timing of stimulus actions (shaded bars) and corresponding utterances. This graph presents the most typical scenario, in which the previous utterance was still in progress when the target action began (92% of trials). The times shown on this graph represent the actual manipulated action durations, and average duration (or estimated durations) of utterances and pauses. The bottom panel illustrates the different relations that obtain from the utterance timing in relation to the action timing. In addition to the two latency components discussed above (planning silence and overlap), on 8% of the trials the previous sentence ended before the target action began, leading to a "waiting gap".
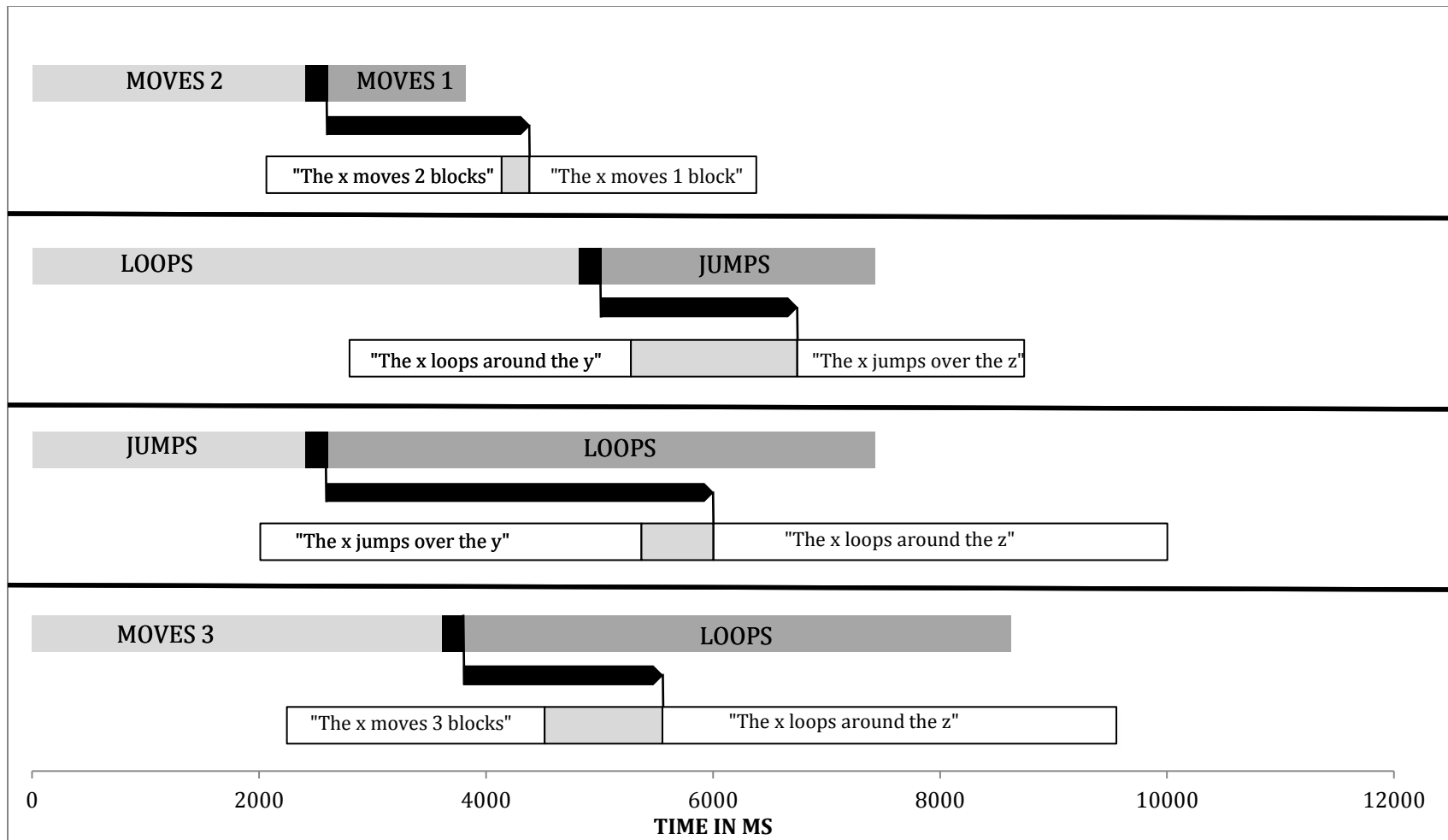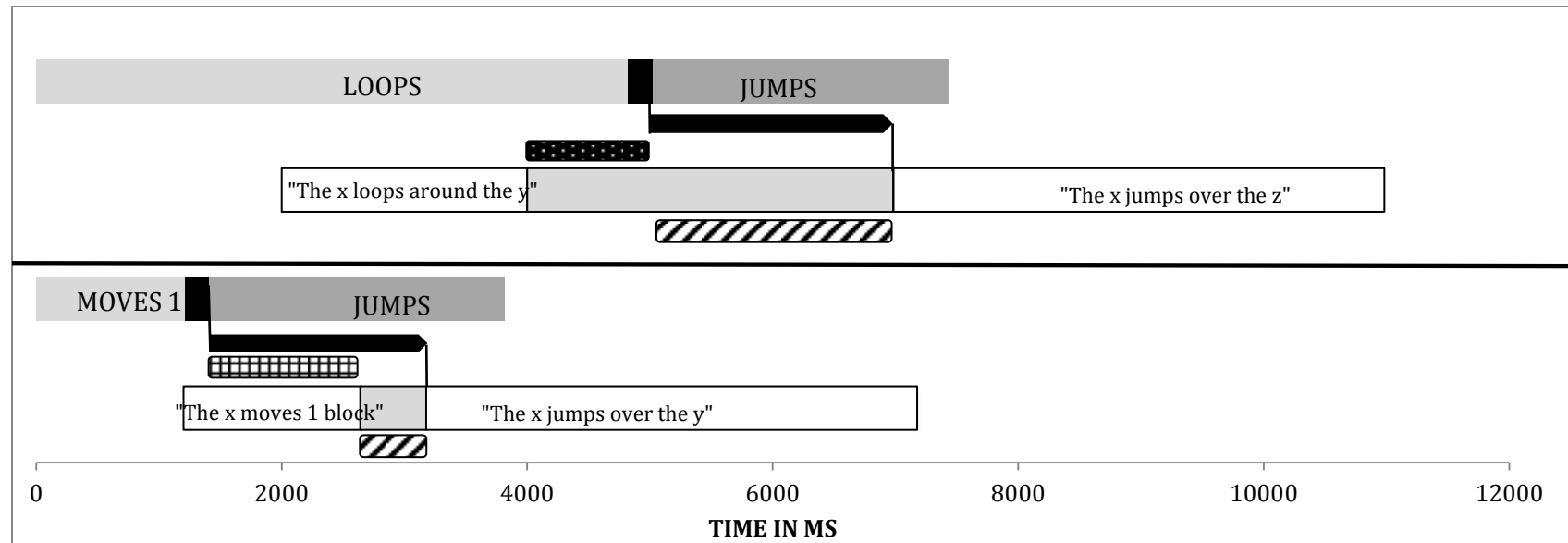
Figure 3. Illustration of timing variables. Top panel illustrates the typical timing of consecutive action descriptions in this database. For each pair, the shaded bars illustrate the duration of the stimulus actions, and the black separator illustrates the 200 msec between actions. These times are fixed. The open bars illustrate average onset and duration of the description for each movement, and the light shaded bar represents the average inter-utterance separation (i.e. "planning silence") for the shown verb pairs[3]. Note that the duration of utterance 2 was not coded, and is only estimated here. Bottom panel illustrates gap, planning silence, and overlap measures (utterance times here are examples and not averages).

---

[3] On the majority of trials (92%), "planning silence" is equivalent to the time between utterances. On the 8% of trials where the prior utterance ended before the action began, "planning silence" is only the time between action onset and target utterance.

4. Connector words. The use of the connector words *and* and *then* was identified from the transcription, and discourse connectors were coded as a binary variable (trials that contained *and*, *then*, or both, vs. those that did not). No other connector words were used.

RESULTS AND DISCUSSION

General Performance

This section describes performance on the entire set of utterances, not just those included in our analyses. A total of 198 trials (out of 2948) were excluded, ranging from 1% to 28% for individual participants. Reasons for exclusion were: 1) Perceptual or memory error (n=10)[4]; 2) merging two consecutive actions into a single description (n=11), 3) a major repair (n=1), 4) technical problems (n=9), 5) missed trials (n=88), or 6) trials following a missed trial, which changed the discourse context (n=79). Participants in both groups were able to do the task, successfully describing between 85% and 100% of trials not excluded for other reasons. Average success rate was 96% in the sham condition, and 98% in the anodal condition.

A preliminary examination of reference form shows that participants in both conditions were heavily influenced by the discourse context, as expected. Here and throughout, the analysis for N1 excludes trials with an error or correction on N1 (N=87) and the analysis for N2 excludes trials with an error or correction on N2 (N=40). No participant used a pronoun or zero when the referent was new, and very few reduced forms were used for N2 (5% for sham, 6% for anodal). For N1, participants frequently used reduced forms on trials where the same object had moved

---

4 As the participant was performing the task, the experimenter identified cases as perceptual/memory error if the participant fell behind significantly and began describing the action after the entire action was over, and the participant was uncertain about what the event was.

on the previous trial: 78% (sham) and 84% (anodal). By contrast, participants almost never

(<1%) used reduced forms for N1 when the target object had been mentioned in nonsubject

position on the previous trial.

Thus, the rest of the analyses in this paper will focus on the references to N1 when the

same object had moved on the previous trial (n=690), which is the only condition with

substantial variation in reference form. Each subject saw 36 items where the same object moved

as on the previous trial. Out of the total 792 possible items in this sample (36 items x 22 subjects),

66 (17%) were excluded for the sham group, and 36 (9%) for the anodal group, for the reasons

listed above. Table 2 illustrates the properties of the 690 trials in our database. This dataset is

available at https://arnoldlab.web.unc.edu/publications/supporting-materials/arnold-nozari-2017/.

Table 2. Properties of the items in our database. Short actions were those with durations 1200-
2410; Long actions were those with durations 3610-4820.

	SHAM		ANODAL	
	Short actions	Long actions	Short actions	Long actions
Average latency	2013	2705	1953	2657
% disfluency	31%	42%	28%	41%
% and/then	60%	69%	71%	81%
verb flash N	16		19	
verb jump N	93		95	
verb loop N		100		116
verb move N	101	20	108	22
Total N	210	120	222	138

This project took an analytical approach that combined experimental elicitation of speech

with statistical methods used for corpus analysis. Our dataset consisted of 690 transcribed and

recorded utterances from the experiment: 330 in sham (74 descriptive NPs) and 360 (56

descriptive NPs) in anodal. We first examine the responses of our sham participants for the

purpose of testing our broadest question: Does the timing of utterance production influence reference form choice? In doing so, we tested the three hypotheses (discourse connectivity, interference, and disfluency discussed earlier). We then examined the effect of stimulation by analyzing reference form choice for the anodal group. We first examined the anodal group alone, testing the role of our four timing predictors. We then compared the anodal and sham groups, by using the Sham Timing Model and adding stimulation as a predictor (both main effect and interactions with other predictors). Finally, we considered possible explanations for our stimulation effects by examining connectors and disfluency.

General statistical approach. The primary analyses examined reference form. We grouped all reduced expressions (pronouns and zeros) together, and compared them with the production of descriptive NPs (e.g., the pink pentagon). This binary dependent variable was analyzed using SAS proc glimmix[5], with a binomial distribution and a logit link. Unless otherwise specified, all models included random intercepts for subject and item (where item represents the particular trial), and random slopes for stimulation condition by items, and for trial-level predictors by subjects. We did not include random slopes for trial-level predictors by items. Even though some of these predictors were not fixed for each item (e.g., onset latency, previous onset latency, presence of disfluency or connectors), the random slopes for these predictors were estimated to be zero in the vast majority of models reported here.[6] For slopes by subjects, if the model estimated a random slope to be zero, it was removed from the model, and the random effects structure is reported.

---

[5] In analyses of latency (a quantitative variable), we followed a similar procedure using SAS proc mixed.

[6] In a very few models, it would have been possible to include a non-zero random slope by items. However, these effects have no impact on the pattern of results, so for consistency they are not included in the reported models.

Because our predictors of interest (IV-1,2,3,4, connector use and disfluency measures) were potentially intercorrelated, we took caution to consider potential collinearity amongst the predictors in each of the models discussed. All predictors were centered, as a first step to reduce collinearity (Jaeger, 2011). We then examined the bivariate correlations amongst all the predictors used in our models. None of these correlations had an absolute value greater than 0.377 (except where otherwise noted), suggesting that collinearity is not a major concern in our models.

Sham analyses: How does the timecourse of message planning affect reference form?

So as not to overfit the model, we first tested each of the four timing predictors (IV-1-4) in separate models. We then selected those that contributed to the model at a level of $t > 1.5$ for a combined model of timing predictors. The significant effects from this model constituted the final Sham Timing Model. We then used this model to examine hypotheses about why the timecourse of utterance production matters, testing the three hypotheses discussed earlier.

The effect of two of the four predictors reached the significance criterion of $t > 1.5$ in separate logit regression models (Table 3). 1) Current latency (IV-1): as predicted, trials with short latencies (indexing limited message preplanning) were less likely to include reduced forms than trials with long latencies (Fig. 4). 2) Previous-action-duration (IV-3): Trials following long actions were considerably less likely to use reduced expressions (55%) than trials following short actions (87%), regardless of the latency of the previous utterance (Fig. 5).

Current utterance timing effects (sham)

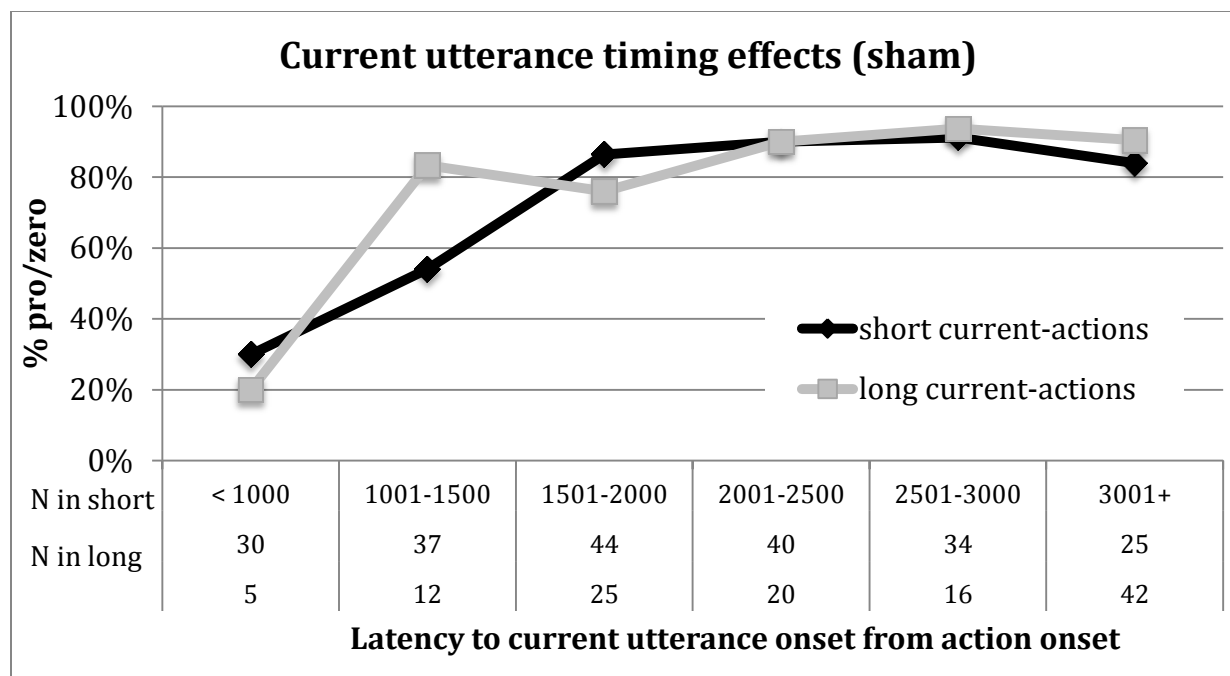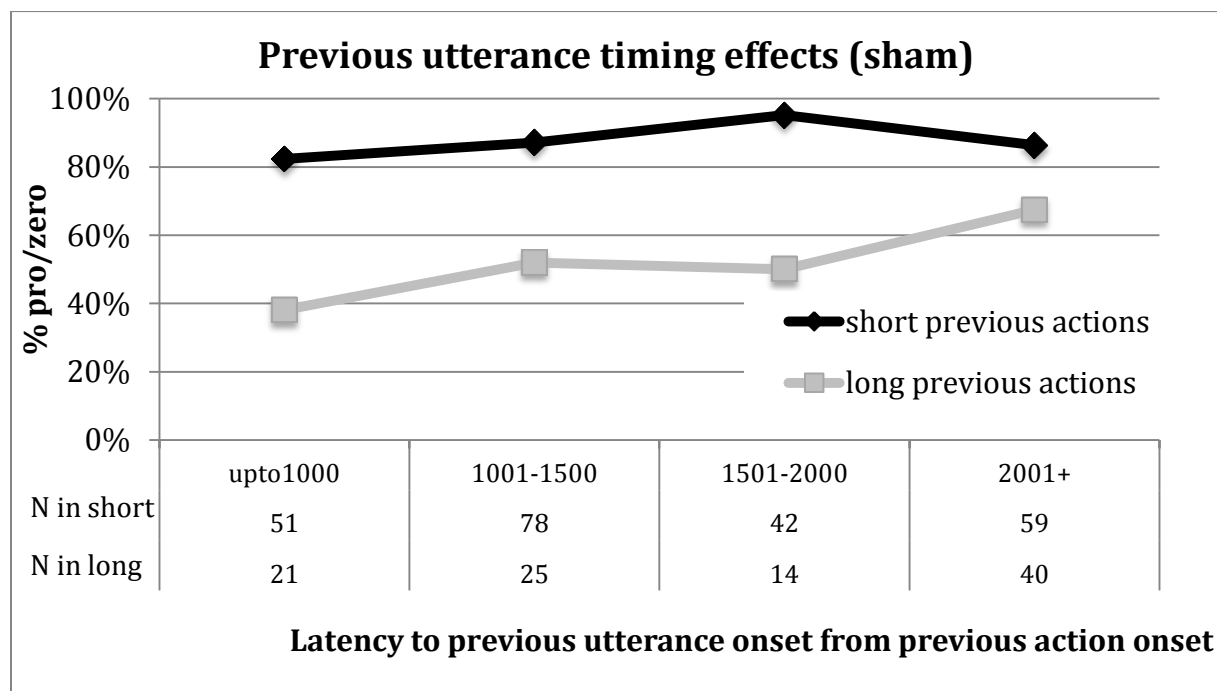| N in short	< 1000	1001-1500	1501-2000	2001-2500	2501-3000	3001+
	30	37	44	40	34	25
N in long	5	12	25	20	16	42

Latency to current utterance onset from action onset

Figure 4. Current Utterance Timing effects: the rate of using reduced forms (pronouns or zeros) as a function of the latency to begin speaking, and the duration of the target action (short durations = 1200-2400 ms; long durations = 3610-4820 ms).

Previous utterance timing effects (sham)

	upto1000	1001-1500	1501-2000	2001+
N in short	51	78	42	59
N in long	21	25	14	40

Latency to previous utterance onset from previous action onset

Figure 5. Previous Utterance Timing effects: the rate of using reduced forms (pronouns or zeros) as a function of the latency to begin speaking on the previous trial, and the duration of the previous target action (short durations = 1200-2400 ms; long durations = 3610-4820 ms).

The two predictors that contributed significantly in their independent models were combined in the final Sham Timing Model, which revealed a main effect of latency (β =.74 (.27), t=2.7, p=.02), and a main effect of previous action durations (β = -0.61(.26), t= -2.3, p=.04). This illustrates that both current-latency and previous-action-duration are independently related to reference form choice, despite the fact that the two predictors are correlated (R =-0.38), such that trials with short previous-action-durations tended to have longer current-latencies (average 2588 ms) than the trials with long previous-action-durations (average 1521). Current-latency was also correlated with IV-2 previous-latency (R = .29) and IV-4 current action durations (R = 0.30), but these predictors did not reliably affect pronoun use.

Table 3. Model results: Effects of utterance timing (in seconds, centered) on reference form. Each predictor is modeled separately. Each model included a random slope by participants (except the Latency of previous utterance model, where the random slope was estimated to be 0).

	Predictor	Description	Estimate (Standard error)	t	p
Current utterance timing	Latency to speak	More pronouns/zeros for longer latencies	1.09 (.33)	3.26	0.008
	Action durations	No effect	0.14(.22)	0.63	0.54
Previous utterance timing	Latency of previous utterance	No effect	0.28 (.20)	1.45	0.15
	Previous action durations	More pronouns/zeros for shorter previous action durations	-0.85 (.26)	-3.27	0.007

In summary, we found that reduced referential expressions were predicted by two timing measures. First, pronouns and zeros occurred more often on trials with long latencies. Second, reduced forms were more common when the previous trial had a short action than when it had a long one. We consider several explanations for these effects below.

Testing Hypothesis 1: Discourse Connectivity

Reduced referential expressions are linguistic devices that enable the speaker to signal a connection between one utterance and the previous context. In our analysis, the previous discourse always supported pronoun use, but we hypothesized that there would be inter-trial variation in the speaker's reliance on this context. When the speaker was better able to conceptualize the relation between the target event and the context, the speaker should be more likely to select a reduced expression. Prediction 1 of this hypothesis was that longer current-latencies and shorter previous-action-durations would support greater message pre-planning, and better consideration of the prior context. As shown above, we saw both of these effects. Current-latency reflected, in part, the degree of overlap between the previous message and the current message, where longer latencies reflected greater overlap of the two conceptual messages and hence greater conceptual coherence.

Similarly, shorter previous-action-durations leave less room for a conceptual gap between the two utterances and also predict greater coherence. Previous evidence has shown that conceptual breaks in the content of a discourse lead to the use of nonreduced forms (McCoy & Strube, 1999; Vonk Hustinx, & Simons, 1992). Thus, a descriptive NP after a gap in the action

may signal the start of a new discourse segment. In our task, a long stimulus action introduced a gap because the action took longer than the speaker needed to plan the utterance.

Further evidence of the conceptual gap between events comes from the fact that we also observed longer silences between utterances when the previous event was long. Trials following long actions tended to have relatively longer breaks between utterances (M = 823 ms) than trials following short actions (M = 390 ms). In some cases these were the result of the speaker finishing the description of the previous trial before the target action had started, requiring the speaker to wait for the action. We identified the trials that required this kind of "waiting gap", classified as those trials on which there was a 50 ms pause or longer between the end of the previous utterance and the onset of the target action. On this criterion, almost none of the trials with short previous actions had a waiting gap (only 1 out of 230 trials), but 20 out of 100 trials with long previous actions did. When we modeled the effect of the waiting gap by itself, it had a significant influence on the tendency to use nonreduced forms (β = -1.22(.42), t= -2.94, p = .004). However, the waiting gap was highly correlated with previous action durations (p = .43), and did not significantly contribute to the Sham Timing model (p < .4). Nevertheless, it supports the conclusion that previous action durations were important because they constrained the continuity of the action input, such that long actions created a gap between planning consecutive utterances.

In summary, the results supported the predictions of the discourse connectivity hypotheses regarding the current-latency (prediction 1) and previous-action-duration (prediction 3). By contrast, the previous-latency and current-action-duration did not account for those effects.

Prediction 5 of the discourse connectivity hypothesis was that current-latency and the previous-action-duration should also predict the production of connector words, and they did. As shown in Figure 6 (right panel), the rate of connector production increased for trials with longer

latencies ($\beta = 0.90(0.24)$, $t = 3.76$, $p = .01$). In addition, connectors were more frequent on trials with short previous-action-durations (71%) than long (47%; $\beta = -0.53$ (0.24), $t = -2.26$, $p = .04$). However, when current-latency and previous-action-durations were entered as predictors in the same model, only current-latency was significant, given the correlation between the two predictors.
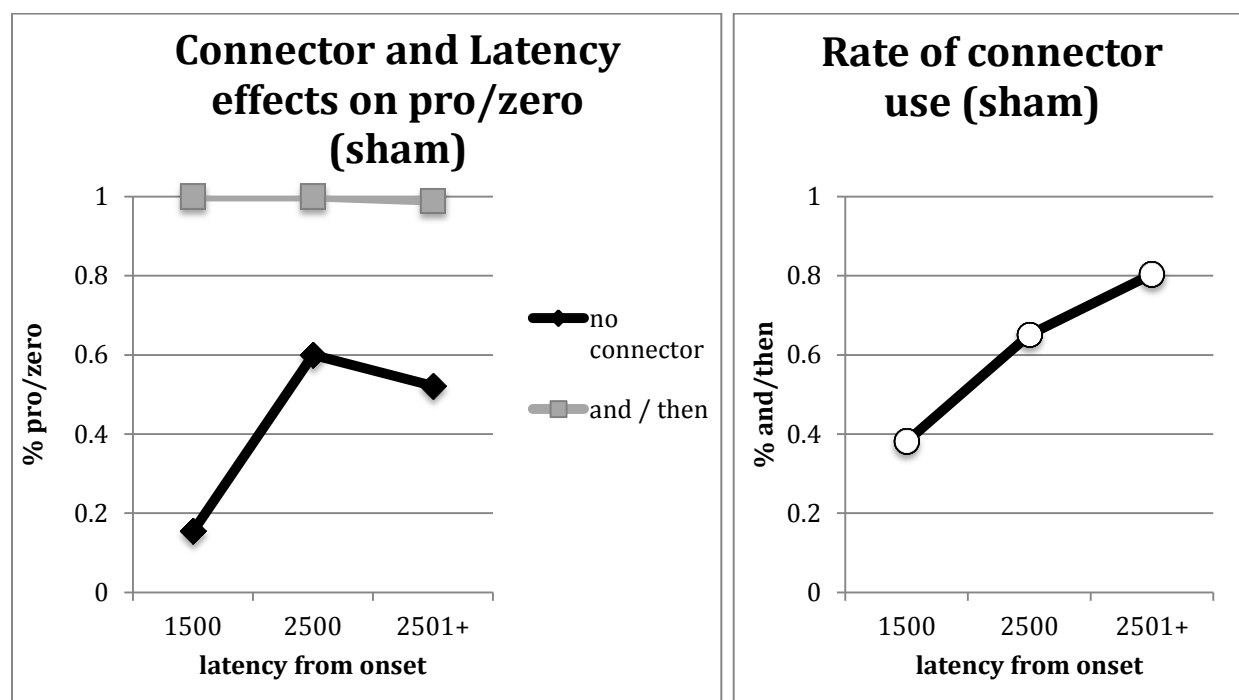

Figure 6. The relation between latency, connectors, and reference form in sham participants. Left panel: reduced expressions are more frequent for trials with connectors at all latencies. Right panel: the rate of connectors increases for trials with longer current-latencies

A parallel prediction was that the connector use itself should predict reference form, which we also observed (see Figure 6 left panel). In fact, the presence of an overt connector was by far the best predictor of the use of a reduced expression: when speaker said *and* or *then,* they

nearly always used a reduced expression as well (left panel). We tested this pattern by adding

connector use as a binary predictor to the Sham Timing Model (random intercepts only, the

slopes by subjects were estimated to be zero). In this model, connector use was a highly

significant predictor (β = 6.23(1.13), t=5.51, p<.0001), as was previous-action duration (β = pl-

0.80(.28), t=-2.83, p=.008). However, the current-latency effect disappeared (β = 0.36(.32),

t=1.12, p=0.26).

Testing alternate hypotheses

The interference hypothesis. Recall that the predictions of the interference hypothesis were the

opposite of those made by the discourse connectivity hypothesis. This is because each hypothesis

makes different predictions about how speech is influenced by the overlap between the previous

and the current utterance. While the discourse connectivity hypothesis suggests that it increases

conceptual coherence, the interference hypothesis suggests that overlap may impair the linguistic

operations underlying the choice of the appropriate reference form. The findings of the analyses

reported above do not support the predictions of the interference hypothesis, since we saw

instead that greater overlap supported the use of reduced forms and connectors.

To test this hypothesis more directly, we measured "planning silence", defined as the

time after the previous utterance had ended and the action had begun, up until utterance onset.

This variable directly measures the period during which interference is at its minimum, because

all the processes related to the previous utterance have finished. We found that planning silence

had no effect on reference form, either when modeled as the only predictor (p = .4) or when

added to the Sham Timing Model (β = -0.0003 (.0004), t = -0.85, p = .40). Together with the

findings of the prior analyses, the planning silence analyses refute the interference model as a

plausible explanation of reduced form production.

The disfluency hypothesis. The pattern of our results was against the interference hypothesis. However, before we take this pattern as conclusive support for the discourse connectivity hypothesis, we must refute disfluency as a mediating variable. Short latencies can lead to both inappropriate referring forms and disfluent repairs (Brown-Schmidt & Tanenhaus, 2006), and disfluency itself is correlated with over-explicit referential forms (Arnold & Griffin, 2007; Arnold et al., 2009). This raises the possibility that the latency effect is mediated by disfluency – i.e., that short current-latencies led to conditions that encouraged disfluent speech, and the disfluency itself triggered the use of overexplicit forms. Our question is whether the presence of disfluency anywhere in the utterance discourages pronoun use, since difficulty planning one part of an utterance can lead to disfluency in other parts of the utterance.

Consistent with the known correlation between disfluency and message planning, we found that speakers tended to be disfluent on trials with short current-latencies (Figure 7, right panel). In a model of the fluency of each trial as a binary dependent measure (fluent vs. disfluent), latency significantly predicted the presence of disfluent elements ($\beta = -0.465(.16)$, t= -2.94, p = .004), and the two predictors were correlated (r = -0.18, p = .001). However, while disfluency by itself was a marginally significant predictor of reference form ($\beta = -0.83(.43)$, t=-1.92, p=.09), it was not significant when it was added to the Sham Timing Model (p=.38), where the only significant predictors were current-latency ($\beta = .68(.28)$, t=2.47, p=.04) and previous-action-duration ($\beta = -0.60(.26)$, t= -2.30, p=.04). As shown in Figure 7 (left panel), the current-latency effect holds for both fluent and disfluent trials, despite the small numeric trend for reduced expressions to be more frequent in fluent than disfluent trials. This finding allows us to reject the disfluency hypothesis as the explanation for the pattern of reduced forms.

Disfluency and Latency effects (sham)
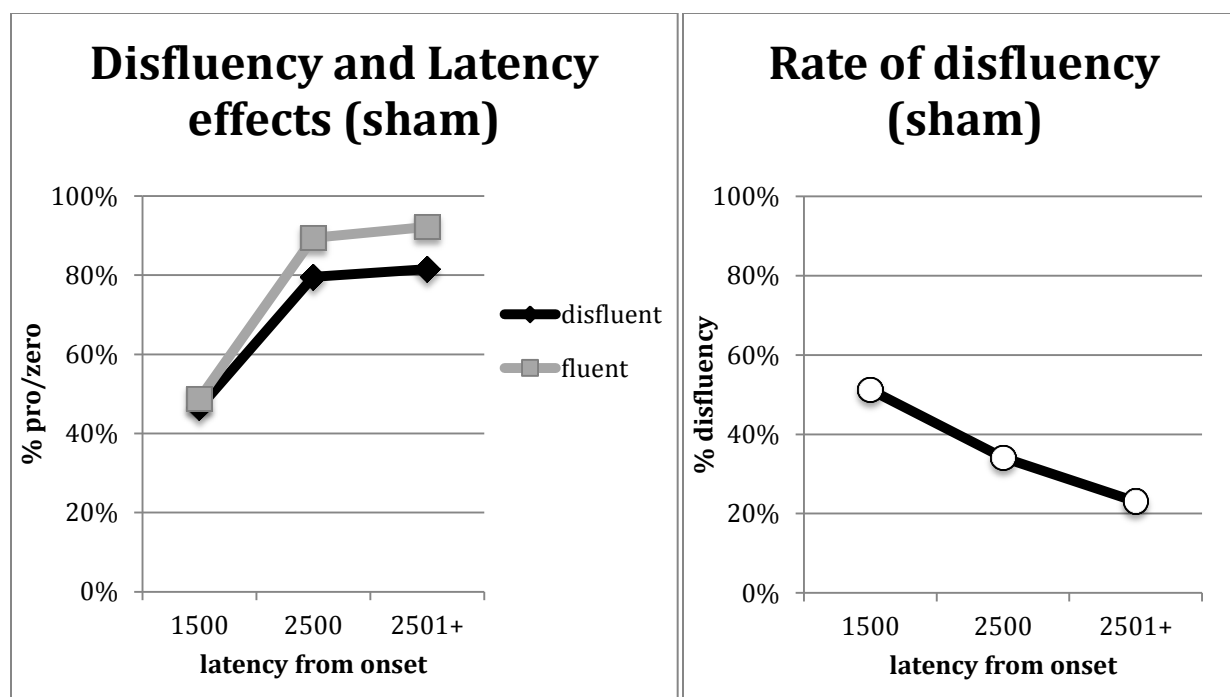
Rate of disfluency (sham)

Figure 7. The effect of disfluency on reference form in sham participants. Left panel: reduced expressions are more frequent for trials with longer current-latencies, for both fluent trials (grey) and disfluent trials (black). Right panel: the rate of disfluency diminishes for trials with longer current-latencies.

Summary of Sham performance

In summary, the timecourse of utterance production had strong effects on the use of reduced referential expressions. We observed two effects. First, longer current-latencies supported the production of reduced forms, while shorter current-latencies increased the use of full descriptions. This effect was strongly tied to variation in discourse connectivity, and once we added connector use as a predictor in the model, the current-latency effect disappeared. Second, the duration of the previous action mattered. When the previous action was short, speakers used more reduced expressions than when the previous action was long. This is likely due to the impact of action timing on discourse connectivity – long actions imposed gaps between planning consecutive utterances, decreasing the sense that the actions were related. Collectively, these results supported the discourse connectivity hypothesis.

Comparison with anodal stimulation

The next question is how performance in our task is affected by changes in the neural tissue that mediates utterance planning, that is, left PFC. Our analyses of the timing variables that contributed to the generation of reduced forms in the previous section allowed for a detailed inspection of the effect of PFC stimulation on those variables. Our first question was whether stimulation would affect reference form. If so, we further asked whether it did so by changing discourse connectivity, by changing the timecourse of planning itself, or by affecting the fluency of utterance production.

We began by examining the effects of our timing measures on the anodal group alone. We utilized a model similar to the basic Sham timing model that included both timing measures (current-latency and previous-duration). As for the sham group, there was a greater use of reduced expressions for trials following short actions than those following long actions (see Figure 9). However, there was no effect of current-latency (see Figure 8). This emerged as a main effect of the previous-action-duration ($\beta = -0.73$ (.22), t=-3.31, p=.003), but no effect of current-latency on reference form ($\beta = -0.02$(.23), t=-0.07, p=.94).
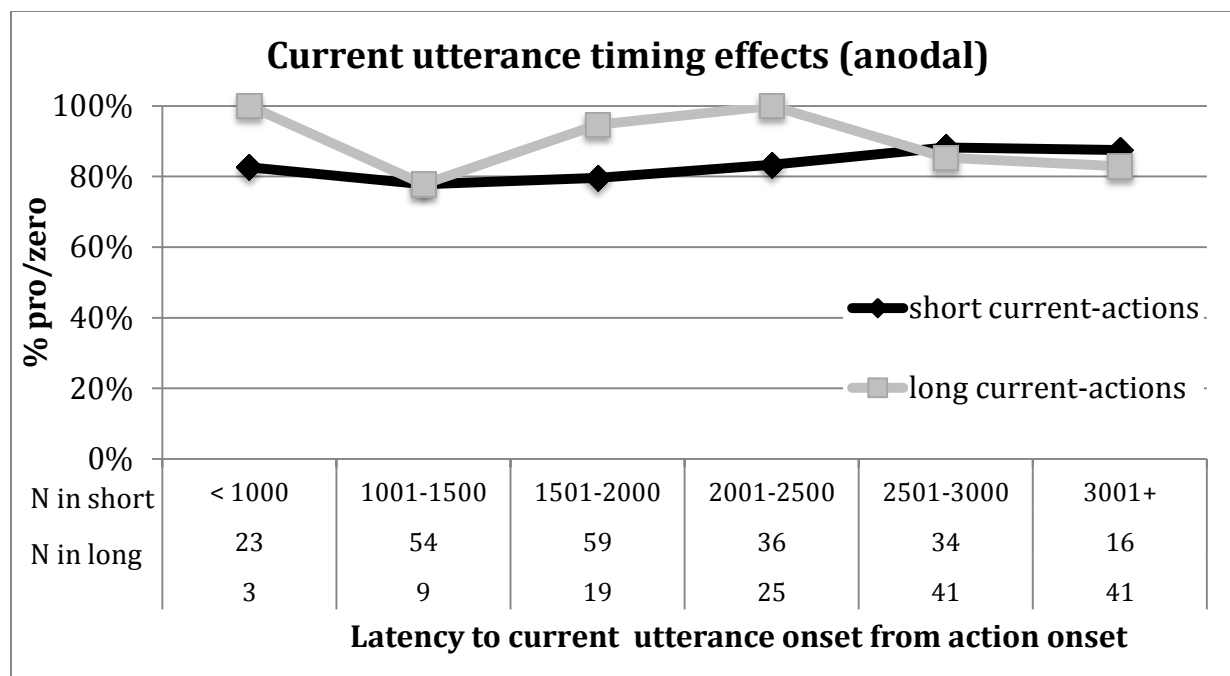
Figure 8. Current Utterance Timing effects in the anodal group: the rate of using reduced forms (pronouns or zeros) as a function of the current-latency to begin speaking, and the current-action-duration (short durations = 1200-2400 ms; long durations = 3610-4820 ms).
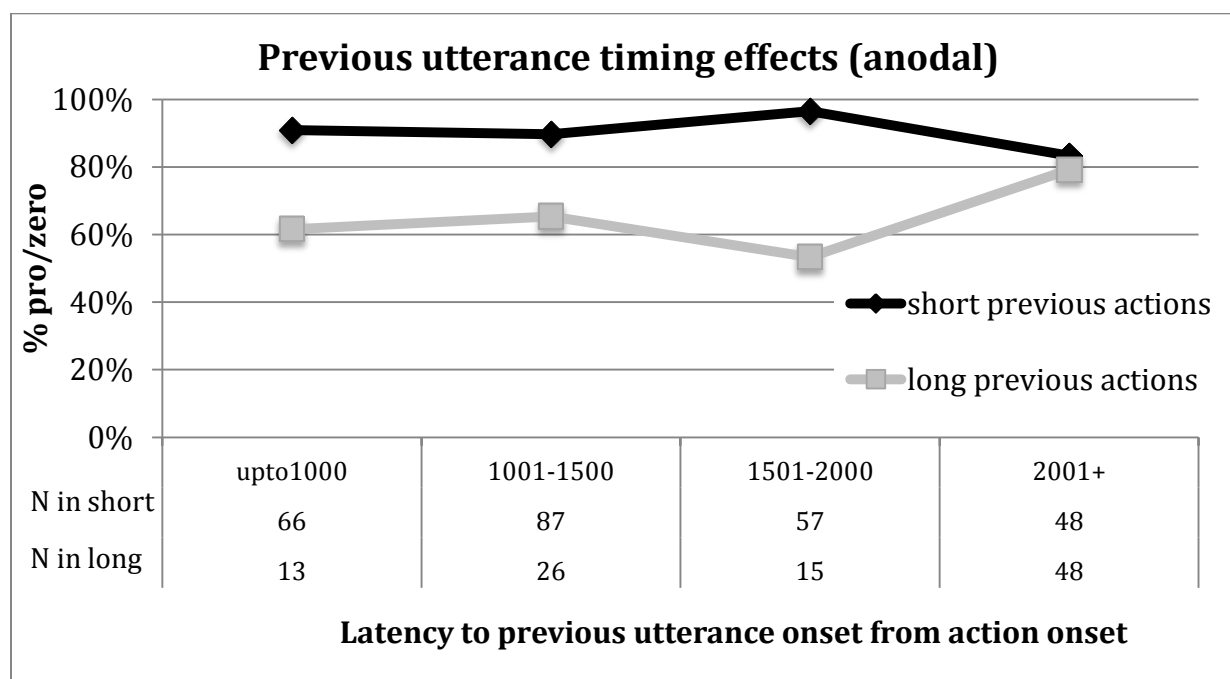

Figure 9. Previous Utterance Timing effects in the anodal group: the rate of using reduced forms (pronouns or zeros) as a function of the previous-latency and the previous-action-duration (short durations = 1200-2400 ms; long durations = 3610-4820 ms).

We examined the role of stimulation by comparing anodal and sham groups with our basic sham timing model, adding stimulation condition and the interaction between stimulation and each predictor. As table 4 shows, there was a main effect of current-latency, and a main effect of previous duration, as well as an interaction between anodal stimulation and current-latency[7].

Table 4. Results of the logit regression analyzing the effect of stimulation and timing measures on reference form production.

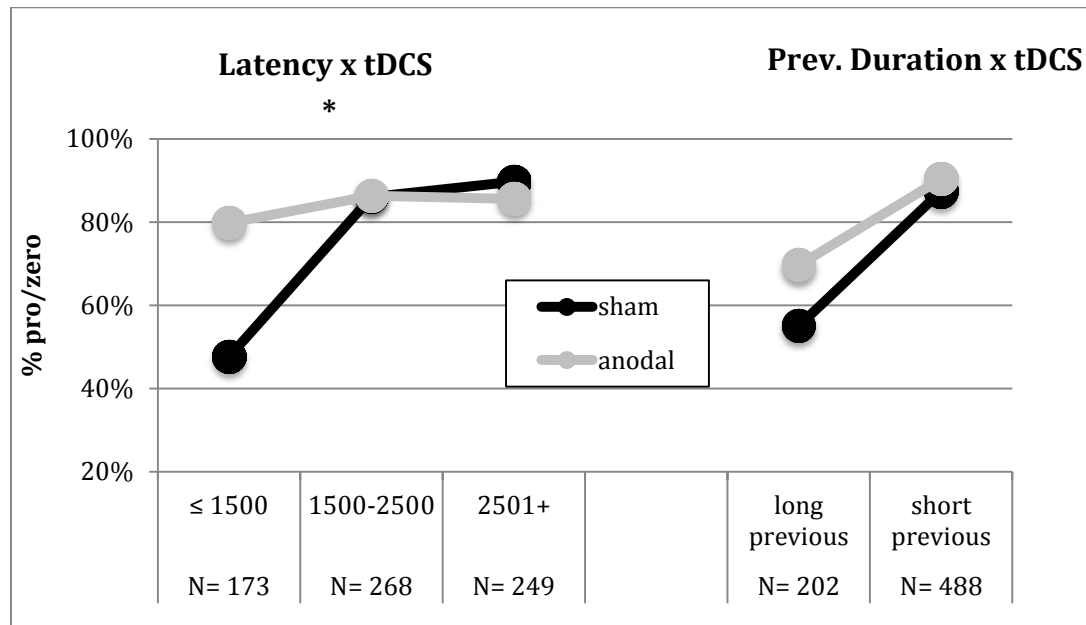| Predictor	Estimate (Error)	t	P	Sig.
Anodal stimulation	0.47 (0.72)	0.66	0.519	
Current onset latency (sec.) (IV-1)	0.49 (0.18)	2.65	0.007	**
Previous action duration (IV-3)	-0.69 (0.21)	-3.24	0.003	**
Anodal x Latency	-0.91 (0.31)	-2.93	0.004	**
Anodal x previous duration	-0.2 (0.26)	-0.74	0.471	

Figure 10. Illustration of the effects of tDCS stimulation, current-latency and previous-action-duration on reference form. Long previous trials are those with action durations 3610 or 4820 ms; short action trials are those with durations in the 1200-2410 ms range.

---

[7] This model had random intercepts for subject and item, and a random slope for subjects by previous-action-durations, but the random slopes for current-latency (by subjects) and stimulation (by items) were both estimated to be zero.

In sum, we found that stimulation did affect reference form, but not across the board. Instead, it moderated the effect of current-latency that we observed for the sham group. For utterances with long latencies, both sham and anodal participants used reduced forms frequently. However, when utterance latencies were short, stimulation made a difference: sham participants used more explicit expressions, while participants under stimulation used reduced forms in a discourse-appropriate way (see Figure 10).

We next considered the mechanism underlying our stimulation effect. One possibility was that stimulation might affect the timing of utterance articulation itself. We tested this in a model using SAS proc mixed, with the log of current-latency as the dependent measure, and stimulation as a predictor. Yet stimulation had no reliable effect on current-latency itself (sham: $M = 2268$ ms; anodal: $M = 2221$ ms, t=0.26, p=.80).

Another possibility was that stimulation would affect the speaker's use of the discourse context in multiple ways, including the choice of reference form. We tested this by examining the effect of stimulation on discourse connectors. As Figure 11 (left panel) shows, anodal stimulation led to an increase in the production of connectors overall, especially for trials with short current-latencies. In a model with connector use as a binary dependent variable, we found a significant effect of current-latency ($\beta = .41(.15)$, t=2.64, p=.01), and an interaction between stimulation and current-latency ($\beta = -0.63(.24)$, t= -2.59, p = .01). Separate analyses by group revealed that current-latency predicted connector use for the sham group ($\beta = 0.90 (.24)$, t= 3.76, p = .01), but not the anodal group ($\beta = 0.06(.18)$, t= 0.32, p = .75). Just as for the sham group, trials with overt connectors almost always used a reduced form (94%), but trials without overt connectors did so less frequently (59%). Thus, stimulation appeared to increase the speaker's sensitivity to the discourse context, even for utterances with short latencies.
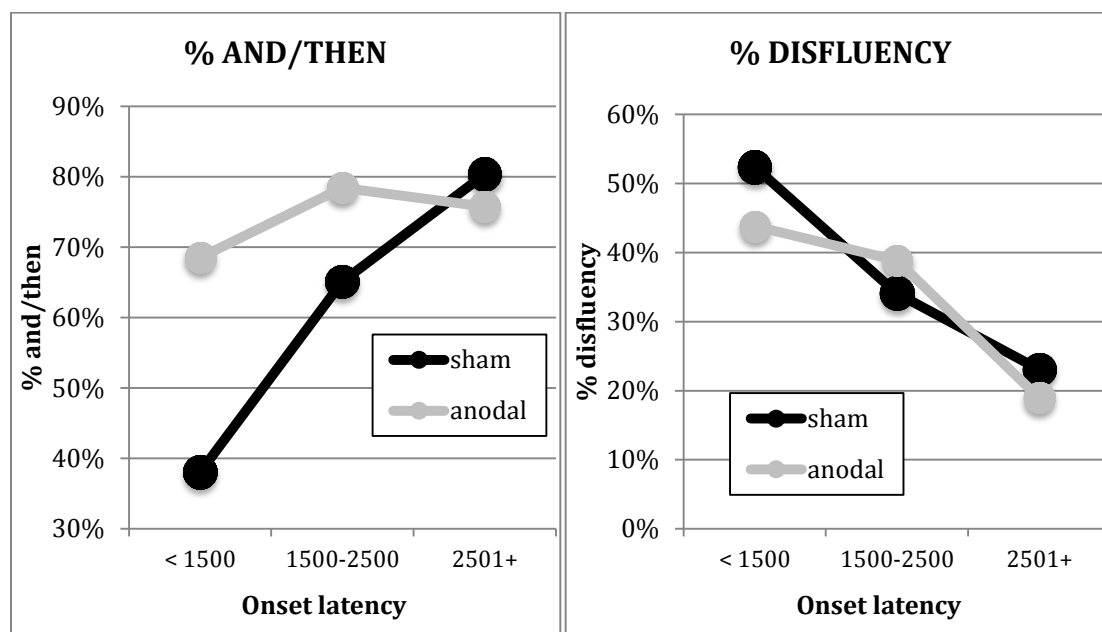
Figure 11. For each group, rate of trials with overt connectors (*and* or *then*; left panel) and rate of disfluent trials (right panel).

On the other hand, figure 9 and the right panel of 10 show that stimulation did not change the effect of previous-action-duration on reference form. For both groups, there were more descriptive NPs for trials with long previous-duration, compared to trials with short previous-action-duration. Similarly, connectors were more common for trials with short previous-action-durations, and this relationship was not affected by stimulation: the rate of connector production was greater for short previous-action-durations (β = -0.60 (0.20), t = -305, p = .005), but there was no effect of either stimulation (p=.33) or the interaction between the two (p=.63).

The dissociation between the effects of latency and previous-action-durations is important in conceptualizing the roles of these two timing factors in reduced form production. Recall that we hypothesized that both are indices of conceptual coherence, but for different

reasons: current-latency reflects planning strategy, where the speaker determines how much overlap there would be between the two utterances. That is, current-latency is more strongly influenced by the speaker's choice (do I want to wait and pre-plan or do I want to start immediately and plan incrementally?) Previous-action-duration, on the other hand, is a property of the event structure itself. When event-conceptualization for the previous event is over much earlier than the beginning of the next event, the conceptual gap may lead speakers to assume a new discourse segment. In this case, the use of an explicit description is the appropriate choice to reflect the structure of the events. Stimulation did not change the previous-action-duration effect. This is consistent with the conclusion that stimulation helps speakers maintain discourse connectivity when it is appropriate to do so, but not when the stimulus timing imposes a break.

Our final question was whether stimulation affected reference form by increasing fluency. However, we did not find strong support for this. Figure 11 (right panel) shows that both groups tended to be more disfluent on trials with short current-latencies more than trials with long latencies. In a model with disfluency as a binary dependent variable, we found a significant effect of current-latency (β = -0.82(.14), t=-5.96, p<.0001). There was no main effect of stimulation (p = .62). Even though stimulation did interact with current-latency (β = -0.54(.23), t=-2.41, p<.02), separate analyses with each group revealed that current-latency had strong effects on disfluency for both groups (sham: β = -0.47(.16), t=-2.94, p=.003; anodal: β = -1.03(.21), t=-4.89, p<.0001). Thus, the effect of stimulation on reference form cannot be attributed entirely to disfluency.

In summary, our data suggest that stimulation increases the likelihood of producing linguistic forms that mark connectivity – both reduced forms, and discourse connectors – and not merely by changing the timecourse of planning or speaker fluency. For the sham group, the

effect of the discourse context was weakened in trials with short latencies, where they used over-specific descriptions and failed to use discourse connectors. For the participants under anodal stimulation, the use of both reduced expressions and discourse connectors was high across the board, regardless of either latency or production difficulty. This suggests that anodal stimulation enabled participants to maintain discourse connectivity even on trials with short latencies.

Reduced forms: pronouns vs. zeros

Our task allowed participants the freedom to choose the phrasing of their utterances, just as they do in natural language use. This meant that they were faced with numerous simultaneous choices. In particular, it permitted them to use two different kinds of reduced expressions: pronouns and zeros. We grouped these together in our analyses, because they are both have similar pragmatic functions, and are primarily used to refer to given and highly accessible information. For this reason, it is impractical to analyze zeros and pronouns independently.

Nevertheless, a qualitative examination of our data suggests that pronouns and zeros fall on a continuum. The conditions that favored reduced expressions also tended to elicit a greater proportion of zeros overall. Figure 12 shows the percentage of zeros and pronouns, and illustrates the numerical tendency for the proportion of zeros to increase for trials with explicit connectors and trials following short actions. This suggests that the effects reported here are particularly impacted by the production of zeros. This is consistent with our conclusion that the observed effects are the result of discourse connectivity, in that elliptical (zero) constructions connect the utterance to the previous one syntactically.
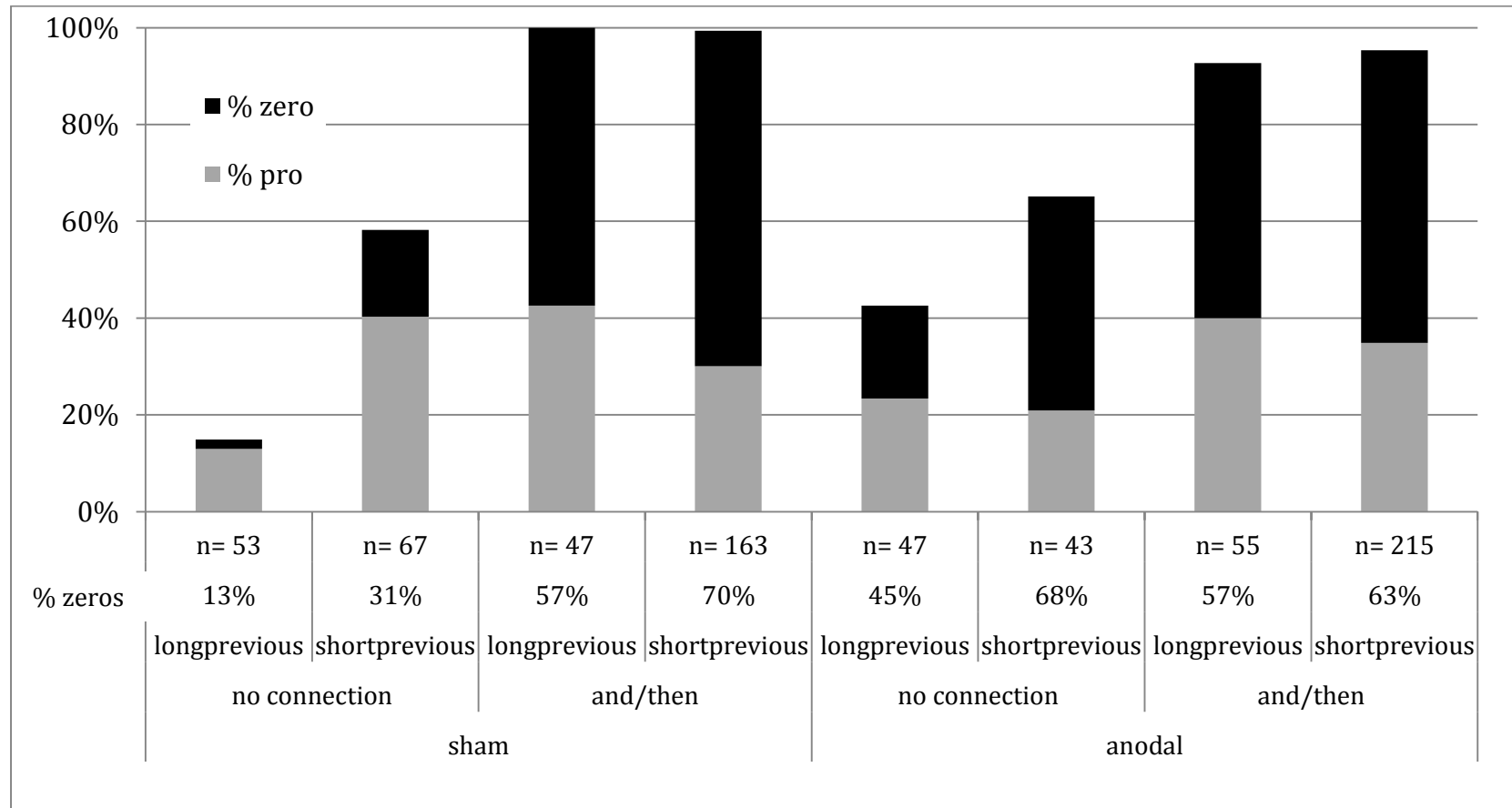
Figure 12. The rate of pronouns, zeros, and both together for each group according to connector use and previous action duration

GENERAL DISCUSSION

The results reported here provide some of the first evidence of how reference form is influenced by the timecourse of utterance planning. We reported two major findings, both of which demonstrate how the timing of utterance initiation can impact variation in reference form.

The first major finding was that the production of pronouns and zeros is influenced by the timing of utterance initiation, as evidenced by two effects. First, for the sham group, longer current-latencies supported the use of linguistic devices that mark discourse cohesion: both connector words (and/then) and reduced forms (pronouns/zeros). By contrast, on trials with short latencies, speakers failed to linguistically mark the connection between that trial and the previous one – despite the fact that our analysis included only trials where pronouns were highly supported, i.e. when the referent was given and highly accessible. Second, when the previous-action-duration was long and it forced a conceptual gap between the two events, speakers used fewer reduced forms, and also fewer overt connectors.

We propose that our data are most consistent with the hypothesis that utterance initiation time affects reference form because it modulates the speaker's ability to represent and/or use the discourse context. In our task, longer current-latencies reflected greater pre-planning, and thus greater overlap between planning the target utterance and speaking the previous one. Likewise, short previous-action-durations led to greater overlap between planning and speaking. We hypothesize that both of these timing variables may have focused the speaker's attention on the connection between the utterances, reinforcing the representation of the fact that the same object did both actions, and leading to linguistic markers of connectivity. By contrast, incremental speaking leads to less overlap, decreasing the strength of the representation of the relation between events. Similarly, we believe that longer previous-action-durations led to a disruption in

conceptual coherence, only this time the disruption was not due to planning strategy, but was

dictated by the properties of the events themselves. This idea in keeping with findings that

speakers mark conceptual or temporal breaks in the discourse with specific referring expressions,

as a way of indicating that a new discourse segment is beginning (Clancy, 1980; Fowler, Levy,

& Brown, 1997; Marslen-Wilson, Levy, & Tyler, 1982; McCoy & Strube, 1999; Vonk, Hustinx,

& Simons, 1992).

An alternate explanation, which we consider less likely, is that the direction of causality

was reversed, such that the choice of referring expression was the driving force behind the

observed timing patterns. The moving object was generally identifiable early in the action. If the

speaker were to immediately select a pronoun or zero, they would be required to delay utterance

onset, because the fluent production of a zero or pronoun requires the speaker to have planned at

least as far as the verb: the zero is only evident once the verb has been uttered, and unstressed

pronouns are typically cliticized onto the verb. By contrast, the selection of a definite NP would

require a fast onset in order to articulate the utterance quickly enough to keep up with the task.

However, this view would predict a consistent relationship between latency and reference form.

This is not what we found, in that the anodal group did not show a significant relationship

between latency and reference form.

Finally, we refuted two alternative hypotheses for the results. The *interference hypothesis*

predicted the opposite of what was found, while *the disfluency hypothesis* predicted a similar

data pattern but attributed the effect to disfluency. Our analysis using disfluency rejected this

hypothesis.

Our second major finding was that stimulation of PFC also affected the speaker's use of

reduced forms, albeit indirectly. While anodal stimulation of PFC did not affect utterance timing

itself, and did not have large effects on disfluency, it did moderate the effect of pre-planning on reference form. For trials with short current-latencies, participants under stimulation managed to produce discourse-appropriate reduced forms, as well as a greater number of trials with discourse connectors than the sham participants. Our findings highlight the importance of PFC for maintaining discourse cohesion via the selection of appropriate linguistic devices.

Interestingly, we found that PFC stimulation affected only one of our two timing variables. Recall that we hypothesized that current-latency and previous-action-duration affect discourse connectivity for different reasons: current-latency reflects the overlap between consecutive messages, while previous-action-duration reflects the properties of the events. We found that PFC stimulation did not affect the stimulus-driven effect (previous-action-durations), and instead was specific to the effect of current-latency, which is partially under the speaker's control.

This finding is broadly consistent with claims that PFC is important for executive function processes generally, and planning specifically (Jurado & Rosselli, 2007; Lehto et al., 2003; Nozari, et al. 2014). Although we do not have direct evidence of the specific PFC mechanisms involved, we can evaluate the hypotheses that PFC stimulation may have influenced 1) working memory, 2) switching between tasks, and 3) interference.

Working memory. The pattern of PFC results may be due to PFC's well-established role in working memory (e.g., Funahashi et al. 1989; Fuster & Alexander, 1971; Kubota & Niki, 1971; see also Curtis & D'Esposito, 2003, and D'Esposito, 2007 for reviews and recent references). Reduced forms like pronouns provide an explicit signal to link to earlier information in the discourse, an operation that must use working memory. An enhancement in working memory under stimulation would be consistent with our finding that anodal stimulation

eliminated the effect of current-latency. We hypothesized that short latencies led to low

conceptual overlap between events in the sham group. Working memory enhancement under

stimulation may bridge the difficulty in linking events in memory, even when speaking

incrementally.

A potential concern with this interpretation is the fact that PFC stimulation did not also

eliminate the previous-action-durations effect. However, recall that this is consistent with our

assumption that the conceptual gap imposed by long previous-action-duration makes speakers

reluctant to view events as related altogether. The stimuli suggest that the events are dissociated,

and speakers represent this dissociation linguistically. Thus, the lower rates of reduced forms as a

function of longer previous-action-durations is not due to failures of cognitive resources and is

insensitive to enhancements of such resources.

An alternative interpretation is that speakers always wish to represent connection in their

discourse, but the longer conceptual gap forced upon them by the longer previous-action-

duration simply makes maintaining the context information in memory more difficult. If so, this

difficulty should be reduced by expanding working memory resources. The fact that we did not

see any interaction between anodal stimulation and previous-action-durations would then

indicate that stimulation did not affect working memory. We consider this explanation less

plausible, given evidence from the literature that working memory is supported by PFC.

Switching Between Tasks. PFC is also known to be involved in shifting or switching

between tasks (e.g., Miyake et al., 2000). If it is the enhancement in this function of PFC that

underlies the current pattern, we would expect improvement specifically in conditions with high

switching demands. Previous-action-duration affects the conceptual connection between events,

but does not change the switching demands. On the other hand, current-latency, as a proxy for

planning mode, is directly related to switching. Incremental planning requires frequent switches between processes at different levels of the production system (plan the first concept, plan the first word, retrieve phonology of the first word, buffer the first word, quickly switch to planning the second word, and so on). While some of these processes can be done in parallel, cognitive bottlenecks in production (e.g., Ferreira & Pashler, 2002) impose a certain degree of serial processing that requires sequencing and switching between tasks. Pre-planning a chunk of message reduces this switching need at least at the level of conceptual planning. This reduces the cognitive load imposed by switching, leaving more resources (e.g., working memory resources) to build conceptual links between events. Thus, we believe that the results are also compatible with the role of PFC in shifting (see Woodard et al., 2016 for data suggesting the particular contribution of shifting to sentence comprehension among other measures of executive control).

Inhibitory control. Finally, PFC has also been strongly implicated in tasks that require inhibitory control (e.g. Aron et al., 2003; Munakata et al., 2011; see Aron et al., 2014 for a review of the most recent evidence). However, we see this explanation as unlikely for our findings, because it requires the presupposition that producing full NPs is a default that must be actively suppressed and inhibited. To our knowledge, there is no empirical evidence in support of this assumption. Even if one assumes that NPs are the default form and must be suppressed in order for reduced form to be produced, it is unclear why better inhibitory control would not just increase pronoun use across the board instead of selectively affecting the current-latency.

In summary, these results provide the first conclusive evidence for the involvement of PFC in production of referential forms, and together with the pattern of behavioral results suggest that this effect is closely related to maintaining conceptual coherence between the two

events, which in turn is relevant to planning strategies adopted by the speaker, as well as

temporal properties of the events.

Conclusion

This study has demonstrated the importance of examining reference form choice in

relation to both the timecourse of utterance planning and the involvement of prefrontal cortex. It

is well known that pronouns and zeros are appropriate only when the referent is given and

accessible in the discourse, but there is substantial variation in how this constraint is applied. The

current results reinforce this core relationship between discourse connectivity and reference form,

suggesting that two conditions promote the speaker's ability to utilize the discourse context: 1)

the lack of temporal breaks in the availability of the conceptual information to be described, and

2) the adoption of a mode of production in which utterances are pre-planned, perhaps in parallel

with uttering earlier utterances. These findings highlight the relevance of production planning

mechanisms to the choice of linguistic forms that promote discourse cohesion. Our tDCS

manipulation supports this conclusion, demonstrating that PFC resources contribute to the

speaker's ability to maintain discourse cohesion even while planning incrementally.

Author note

References

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

Ariel, Mira. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W.
Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 29–
87). Amsterdam: Benjamins.

Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop-
signal inhibition disrupted by damage to right inferior frontal gyrus in humans.
Nature Neuroscience, 6(2), 115-116.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal
cortex: one decade on. *Trends in cognitive sciences*, 18(4), 177-185.

Arnold, J. E. (1998). Reference form and discourse patterns. Ph.D. Dissertation, Stanford
University.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented
processes. *Language And Cognitive Processes, 23*(4), 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., & 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.
doi:10.1016/j.jml.2006.09.007

Arnold, J.E., Bennetto, L., & Diehl, J. J. (2009). Reference Production in Young Speakers with
and without Autism: Effects of Discourse Status and Processing Constraints.
Cognition, 110, 131-146.

Baker, J. M., Rorden, C., & Fridriksson, J. (2010). Using transcranial direct-current

stimulation to treat stroke patients with aphasia. *Stroke*, *41*(6), 1229-1236.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.44)

[Computer program]. Retrieved February 1, 2013, from www.praat.org.

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size:

An investigation of message formulation and utterance planning. *Journal of Memory

and Language*, *54*(4), 592-609.

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

view. In C. N. Li, (Ed.), Subject and topic (pp. 25-56). New York: Academic Press Inc.

Clancy, P. M. (1980). Referential choice in English and Japanese narrative discourse. In W. L.

Chafe (Ed.), *The pear stories: Cognitive, cultural and linguistic aspects of narrative

production* (pp. 127-202). Norwood, NJ: Ablex.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive

Psychology*, 37, 201-242. doi:10.1006/ cogp.1998.0693

Clark, H., & Fox Tree, J. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1),

73-111.

Damasio, A. R. (1992). Aphasia. *New England Journal of Medicine*, *326*(8), 531–539.

Dronkers, N. F., Plaisant, O., Iba-Zizen, M. T., & Cabanis, E. A. (2007). Paul Broca's historic

cases: high resolution MR imaging of the brains of Leborgne and Lelong. *Brain*,

130(5), 1432–1441.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages

of word production. *Journal of Experimental Psychology: Learning, Memory, and

Cognition, 28*(6), 1187.

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. doi:10.1006/jmla.2001.2797.

Fertonani, A., Rosini, S., Cotelli, M., Rossini, P. M., & Miniussi, C. (2010). Naming facilitation

       induced by transcranial direct current stimulation. *Behavioural Brain Research*,

       208(2), 311-318. doi:10.1016/j.bbr.2009.10.030

Flöel, A. (2012). Non-invasive brain stimulation and language processing in the healthy

       brain. *Aphasiology*, *26*(9), 1082-1102. doi:10.1080/02687038.2011.589892.

Fowler, C., Levy, E., & Brown, J. (1997). Reductions of spoken words in certain discourse

       contexts. *Journal of Memory and Language,* 37, 24 – 40.

Fregni, F., Boggio, P. S., Nitsche, M., Bermpohl, F., Antal, A., Feredoes, E., … Paulus, W. (2005).

       Anodal transcranial direct current stimulation of prefrontal cortex enhances

       working memory. *Experimental Brain Research*, *166*(1), 23–30.

Fuster, J.M. (1989). *The Prefrontal Cortex* (Second Edition). New York: Raven Press.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.). The

       psychology of learning and motivation(Vol. 9). New York: Academic Press.

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic*

       continuity in discourse: A quantitative cross-language study (pp. 1–42). Amsterdam:

       The Netherlands.

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and

       articulation of words in speaking. *Psychonomic Bulletin and Review, 10*(3), 603-609.

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

       expressions in discourse. *Language, 69*(2), 274–307.

Hamilton, R. H., Chrysikou, E. G., & Coslett, B. (2011). Mechanisms of aphasia recovery after stroke and the role of noninvasive brain stimulation. *Brain And Language*, *118*(1-2), 40-50. doi:10.1016/j.bandl.2011.02.005

Martin, R. C. & Freedman, M. L. (2001). Short-term retention of lexical- semantic representations: Implications for speech production, *Memory*, 9:4-6, 261-280, DOI: 10.1080/09658210143000173

Martin, R. C. & He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language, 89*, 76–82.

Hartsuiker, R.J. & Barkhuysen, P.N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes 21* (1-3), 181-204

Holland, R., & Crinion, J. (2012). Can tDCS enhance treatment of aphasia after stroke? *Aphasiology*, *26*(9), 1169–1191.

Iyer, M. B., Mattu, U., Grafman, J., Lomarev, M., Sato, S., & Wassermann, E. M. (2005). Safety and cognitive effect of frontal DC brain polarization in healthy individuals. *Neurology*, *64*(5), 872–875.

Jaeger, T. F. (2011). Corpus-based Research on Language Production: Information Density and Reducible Subject Relatives. In Benders, E. M. and Arnold, J. E. (eds): *Language from a Cognitive Perspective: Grammar, Usage, and Processing. Studies in honor of Tom Wasow*, 161-197. Stanford: CSLI Publications.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology review*, *17*(3), 213-233.

Kessler, S. K., Turkeltaub, P. E., Benson, J. G., & Hamilton, R. H. (2012). Differences in the experience of active and sham transcranial direct current stimulation. *Brain stimulation*, *5*(2), 155-162.

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language, 66,* 143–162.

Lehto, J. E., Juujärvi, P., Kooistra, L., and Pulkkinen, L. (2003). Dimensions of executive functioning: evidence from children. *British Journal of Developmental Psychology 21*, 59–80. doi: 10.1348/026151003321164627

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

Luria, A. R. (1970). The functional organization of the brain. *Scientific American*, *222*, 66–78.

Luria, A. R. (1973). The frontal lobes and the regulation of behavior. In K.H. Pribram & A.R. Luria (Eds.), *Psychophysiology of the frontal lobes* (pp. 3-26). New York: Academic Press.

Marslen-Wilson, W., Levy, E., & Tyler, L. K. (1982). Producing interpretable discourse: The establishment and maintenance of reference. In R. Jarvella & W. Klein (Eds.), Speech, place, and action: Studies in deixis and gesture (pp. 339-378). New York: Wiley.

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.

Meyer, A. S., Belke, E., Häcker, C., & Mortensen, L. (2007). Use of word length information in utterance planning. Journal of Memory and Language, 57, 210-231. doi:10.1016/j.jml.2006.10.005.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. Cognitive psychology, 41(1), 49-100.

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. Trends in cognitive sciences, 15(10), 453-459.

Nelson, J. T., McKinley, R., Golob, E. J., Warm, J. S., & Parasuraman, R. (2014). Enhancing vigilance in operators with prefrontal cortex transcranial direct current stimulation (tDCS). *Neuroimage*, *85*(Part 3), 909-917. doi:10.1016/j.neuroimage.2012.11.061

Nozari, N., & Thompson-Schill, S. L. (2013). More attention when speaking: Does it help or does it hurt?. *Neuropsychologia*, *51*(13), 2770-2780.

Nozari, N., Woodard, K., & Thompson-Schill, S. L. (2014). Consequences of cathodal stimulation for behavior: when does it help and when does it hurt performance?. *PloS one, 9*(1), e84338.

Nozari, N., Arnold, J. E., & Thompson-Schill, S. L. (2014). The effects of anodal stimulation of the left prefrontal cortex on sentence production. *Brain Stimulation*, doi:10.1016/j.brs.2014.07.035

Ohn, S. H., Park, C. I., Yoo, W. K., Ko, M. H., Choi, K. P., Kim, G. M., ... & Kim, Y. H. (2008). Time-dependent effect of transcranial direct current stimulation on the enhancement of working memory. *Neuroreport*, *19*(1), 43-47.

Robinson, G., Blair, J., & Cipolotti, L. (1998). Dynamic aphasia: an inability to select between competing verbal responses? *Brain*, *121*(1), 77–89.

Schlaug, G., Marchina, S., & Wan, C. Y. (2011). The use of non-invasive brain stimulation techniques to facilitate recovery from post-stroke aphasia. *Neuropsychology Review*, *21*(3), 288–301.

Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *Journal of Memory and Language, 75*, 27-44.

Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. European Journal of Cognitive Psychology, 11, 17–50.

Shalice, T. (1982). Specific impairments of planning. In P. Broadbent & L. Weisknartz (Eds.), The neuropsychology of cognitive function (pp. 199-209), London: The Royal Society.

Stuss, D.T., & Benson, D.F. (1986). The Frontal Lobes. New York: Raven Press.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, 94*(26), 14792–14797.

Van der Meulen, F. F., Meyer, A. S., & Levelt, W. J. M. (2001). Eye movements during the production of nouns and pronouns. *Memory & Cognition*, 29, 512-521.

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, 73, 301–333.

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple

lexical access. *Journal of Experiment Psychology: Learning, Memory, and Cognition, 36,*
423–440.

Wheeldon, L. R. & Lahiri, A. (1997) Prosodic units in speech production. Journal of Memory
and Language 37:356–81.

Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual
differences in executive function and language processing skills in child learners.
Journal of experimental child psychology, 141, 187-209.

Zaehle, T., Sandmann, P., Thorne, J. D., Jäncke, L., & Herrmann, C. S. (2011). Transcranial
direct current stimulation of the prefrontal cortex modulates working memory
performance: combined behavioural and electrophysiological evidence. *BMC
neuroscience*, *12*(1), 2.

Zerkle, S., & Arnold, J. E. (In press). Discourse attention during utterance planning affects
referential form choice. *Linguistics Vanguard*.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space
in the monkey's dorsolateral prefrontal cortex. *Journal of neurophysiology*, *61*(2),
331-349.

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term
memory. *Science*, *173*(3997), 652-654.

Kubota, K., & Niki, H. (1971). Prefrontal cortical unit activity and delayed alternation
performance in monkeys. *Journal of Neurophysiology,* 34(3), 337-347.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical
Transactions of the Royal Society of London B: Biological Sciences*, *362*(1481), 761-
772.

Appendix A. Distribution of predictors in the dataset as a whole (top) and major dependent measures pertaining to linguistic form (bottom).

	SHAM				ANODAL			
	previous action long		previous action short		previous action long		previous action short	
	target action short	target action long	target action short	target action long	target action short	target action long	target action short	target action long
TIMING CHARACTERISTICS								
N per subject	9	1	13	13	9	1	13	13
Total N in dataset after exclusions	89	11	121	109	91	11	131	127
Avg. Latency (ms)	1539	1379	2362	2838	1659	2109	2157	2704
% trials with a waiting gap greater than 50 ms	22%	0%	0%	1%	36%	0%	0%	1%
Overlap between previous sentence and target action (ms)	818	698	2056	2365	542	714	1802	2053
Avg. Planning silence (ms)	721	681	307	473	1117	1395	355	651
VARIATION IN LINGUISTIC FORM								
% And/then	48%	36%	69%	72%	53%	64%	84%	83%
% Reduced forms	54%	64%	88%	86%	68%	82%	92%	89%
% zero	29%	18%	60%	49%	34%	64%	63%	53%
% Disfluency/correction	43%	73%	22%	39%	43%	55%	18%	39%