Discourse Processes, 50:23–51, 2013

ISSN: 0163-853X print/1532-6950 online
DOI: 10.1080/0163853X.2012.727719

Routledge
Taylor & Francis Group

Effects of Conversational Pressures on Speech Planning

Benjamin Swets
Psychology Department
Grand Valley State University

Matthew E. Jacovina and Richard J. Gerrig
Psychology Department
Stony Brook University

In ordinary conversation, speakers experience pressures both to produce utterances suited to particular addressees and to do so with minimal delay. To document the impact of these conversational pressures, our experiment asked participants to produce brief utterances to describe visual displays. We complicated utterance planning by including tangram figures that prohibited easy lexicalization. Participants completed the task in either the presence or absence of an addressee and also under circumstances of natural or explicit time pressure. Results suggested that speakers produce richer utterances with addressees present but that they do so efficiently, without sacrificing planning time. We propose a good-enough view of the language production system: We suggest that, much like the comprehension system, speech planning processes flexibly adapt to external task goals.

INTRODUCTION

When people plan and execute speech, the dual pressures to say what they mean and to do so within a reasonable time course may create different goals with

Matthew E. Jacovina is now a postdoctoral fellow at the School of Education and Social Policy and the Department of Psychology at Northwestern University.

Correspondence concerning this article should be addressed to Benjamin Swets, Psychology Department, Grand Valley State University, 2224 Au Sable Hall, Allendale, MI 49401, USA. E-mail: swetsb@gvsu.edu

regard to planning (Clark & Brennan, 1991; Clark & Carlson, 1982; Clark & Murphy, 1982; Ferreira & Swets, 2002, 2005; Jefferson, 1989; Sacks, Schegloff, & Jefferson, 1974). If a speaker's primary goal is to produce a fully realized utterance for a listener, his or her extended contemplation may lead to awkward conversational gaps. If a speaker's goal is to initiate utterances as quickly as possible, she or he may sacrifice clarity of meaning. The purpose of our project is to examine the way these goals may affect both the time course with which speech is planned and an utterance's content given a particular time course. In doing so, we evaluate the extent to which important tenets of the good-enough approach to language comprehension (Ferreira, Bailey, & Ferraro, 2002) may also apply in language production. The good-enough approach emphasizes the way in which task goals shape the extent to which comprehenders expend time and mental resources in the service of interpreting sentences (Ferreira et al., 2002; Ferreira & Patson, 2007; Swets, Desmet, Clifton, & Ferreira, 2008). We argue, in parallel, that examining the impact of goals on speech planning will help reveal the architecture of the language production system.

Good-Enough Approach

The good-enough approach to language comprehension has claimed that the process of understanding sentences does not yield perfect and veridical representations of sentence meaning but instead is tuned to addressees' particular goals for the task at hand (Ferreira et al., 2002; Sanford & Sturt, 2002; Swets et al., 2008). This approach assumes that representations of sentences can be strategically underspecified because the comprehension system is capacity limited and that the language-processing system adjusts the dispersal of its resources in service to particular goals.

Early evidence for such a system came from data concerning errors in comprehension, such as when people answer the question "How many of each sort of animal did Moses put on the ark?" with "Two" rather than "I believe you're thinking of Noah" (Reder & Kusbit, 1991). Other studies have focused on the task circumstances that elicit differences in the extent to which comprehenders fully analyze, or reanalyze, a sentence. For example, Swets et al. (2008) found that people who read ambiguous sentences committed different resources to resolution of the syntactic ambiguity depending on whether their goal was to answer detailed versus superficial questions about the ambiguities. Such results suggest that, in comprehension, resources such as time, memory, and attention are at a premium and that the system will strategically spend or save those resources in response to goals inherent to the present task.

We propose that a similar analysis can be given with respect to language production. We predict that speakers can strategically manage the way they use the resources at their disposal—namely time and/or mental resources such as attention and working memory—in response to the goals elicited by different external pressures. Before we make more specific predictions, we describe two types of goals that guide language production: speakers' goals to say what they mean and to do so with an efficient time course.

The Goal of Audience Design

Speakers in conversations have as one goal to design utterances that allow particular addressees to recover the meanings they intend (Clark & Carlson, 1982; Clark & Murphy, 1982; Sacks et al., 1974). Consider this excerpt, taken from the CallHome American English corpus of telephone speech collected by the Linguistic Data Consortium (Kingsbury, Strassel, McLemore, & McIntyre, 1997) in which Speaker A explains to Speaker B what their friend "Lawrence" did while in the military:

> A: {inhale} Really that's what Lawrence and one of his friends that's what he did when he was in the service.
> A: because he just uh you know the eh the
> A: you know how the tops of the tanks have those kind of ball bearing things
> B: mhm
> A: He just made sure that those ran right.

When speaker A begins the utterance "because he," it appears that he intends to name what Lawrence did in the service. However, he is unable to find a simple term that satisfactorily captures the nature of Lawrence's work. Instead, Speaker A attempts to give a clear enough description of the work for Speaker B to understand its nature. The description is somewhat long and circular, but Speaker A eventually provides enough rich detail for Speaker B to confirm that he understands (i.e., "mhm"). This example illustrates how speakers' goals exercise pressure in conversation: pressure to provide adequate detail to establish common ground.

One goal of our study is to demonstrate the pressure the goal of audience design has on the content of speech planning. Most studies of planning in language production have not revealed how conversational pressures might influence speech planning because they have used paradigms that exclude any form of addressee or audience—except, perhaps, the experimenter in the room. Such partner-free circumstances have been labeled "monologue" because speakers in such paradigms lack addressees (Pickering & Garrod, 2004). Pickering and Garrod (2004) critiqued the study of language production in monologue by arguing, for example, that dialogue actually simplifies language production compared with monologue. They suggested that dialogue reduces the decision-making processes required of speakers. Such decisions are facilitated by previously used syntactic structures and words spoken by both parties in the dialogue (Bock, 1986). Pickering and Garrod (2004) concluded that because studies of language production fall short on empirical grounds by being limited to observations of speech without addressees, they risk falling short on theoretical grounds by missing key features of the nature of language production.

We find these arguments compelling and suggest that mechanistic theories of speech planning will benefit from examining the precise manner in which different properties of speech tasks involving concrete addressees exert pressure on speakers' planning processes. Prior research firmly in the dialogue-based tradition of psycholinguistics has focused on the effects of addressees (Murfitt & McAllister, 2001) and partner goals (Russell & Schober, 1999) on the content, but not the planning, of speech. Other work has examined more mechanistic aspects of speech planning in the context of conversation (Brown-Schmidt & Tanenhaus, 2006) without directly comparing planning processes with and without addressees. Our project begins to bridge the gap between these two traditions by examining the impact of the mere presence of an addressee on the time course and degree of incrementality of speech planning.

The Goal of Speaking Promptly

We have just suggested that the presence of addressees creates pressure on speakers to serve the goal of audience design. We suggest, additionally, that speakers plan speech in response to the goal to begin speaking with a reasonable time course. In fact, it is commonplace in conversation for a speaker to begin articulating an utterance without having the utterance fully planned. Consider the excerpt below taken from the CallHome corpus. Speaker C has just introduced the topic of discussion, a mutual acquaintance of the two participants. When Speaker D asks where this acquaintance is, Speaker C's answer is quite disfluent:

C: and uh he's coming back down here in April but
D: Where's he at?
C: eh He's He's in uh uh he's in Utah.

In this example, Speaker C appears to have difficulty retrieving "Utah." He might have waited until his memory processes made that information available before he began his utterance. However, Speaker C responded to ordinary conversational pressure not to leave periods of silence (Jefferson, 1989). That pressure to forge ahead prompted Speaker C to plan some of his utterance before he began to speak, but not enough to avoid the disfluency.

Models of language production generally share the assumption that speech planning functions incrementally (Bock & Levelt, 1994; Kempen & Hoenkamp, 1987; Levelt, 1989). Incrementality entails that when planning at, for example, the level of grammatical encoding, speakers do not plan the entire surface structure of an utterance before beginning to articulate the utterance. Many accounts of incrementality assume rigidly small planning units, such as a phrase at a time (Bock & Levelt, 1994; Griffin, 2001; Kempen & Hoenkamp, 1987; Levelt, 1989; Meyer, Sleiderink, & Levelt, 1998; Smith & Wheeldon, 2001). Other research suggests that planning scope (i.e., the amount of the surface structure speakers plan in advance of articulation) at various levels of representation is flexibly incremental: The size of the planning unit is sensitive to local circumstances (Christianson & Ferreira, 2005; Costa & Caramazza, 2002; Damian & Dumay, 2007; Ferreira & Swets, 2002, 2005; Korvorst, Roelofs, & Levelt, 2006; Schriefers & Teruel, 1999; Wagner, Jescheniak, & Schriefers, 2010). For example, speakers under an explicit deadline plan in smaller increments compared with speakers with no such deadline (Ferreira & Swets, 2002, 2005).

In the current experiment, we test the possibility that the mere presence of an addressee might produce implicit time pressure: Speakers might begin speaking sooner and plan less than they would without addressees present because they want to avoid delays in speech to avoid conversational gaps (Clark & Brennan, 1991; Jefferson, 1989). Thus, it is possible that the presence of an addressee will lead speakers to reduce the scope of speech planning. However, we argue that speakers with addressees plan richer utterances in response to the goal of audience design and do so efficiently, causing little impact on the scope of their planning. To make that prediction concrete, we describe the language game in which we asked participants to engage.

Paradigm

To examine the impact of speakers' dual goals, we devised a game that allowed us to measure the impact of addressee presence and time pressure on speakers' object descriptions. Some participants played the game in the presence of an addressee, whereas some played it alone. In the game a Director produced verbal commands based on visual displays shown on a computer (Figure 1). The visual displays contained three objects aligned horizontally across the screen. Participants produced utterances that mentioned the three objects and the direction of movement (as indicated by the arrows). For example, in response to Figure 1 one participant said, "The shoe moves below the black mountain silhouette and the coffee cup moves below the black mountain silhouette." The speakers understood that the purpose of their utterances was to allow Matchers to manipulate items on a grid displayed on the Matchers' own computer. For Directors in the Addressee Present condition, the Matchers were physically present at the time they uttered the commands. Directors in the Addressee Absent

FIGURE 1 Example stimulus from Experiment 1. The target utterance for this display is *The shoe moves below the [tangram description] and cup moves above the [tangram description]* (color figure available online).

condition produced their commands with the expectation that Matchers might listen to their commands at a later point in time (as they did).

In the game most of the objects were colored line drawings of real-world objects such as a shoe and a cup. The remaining objects were tangrams: objects consisting of shapes arranged into a whole that resembles some real-world object. Speakers do not lexicalize these objects as easily as real-world objects (e.g., Clark & Wilkes-Gibbs, 1986; Horton & Gerrig, 2002). We included tangrams to elicit utterances with a broad range of descriptive detail. One speaker referred to the tangram in Figure 1 as a "mountain." Another speaker referred to it as "the black-colored thing that looks like two mountains on the sides and two triangle-shaped things on top."

In addition to addressee presence, we also manipulated the presence of a deadline to begin speaking. Half the Directors in our experiment produced their utterances under explicit time pressure: We required them to begin speaking in 2.5 s. This explicit deadline echoes past research with monologue and, thus, gives us the opportunity to determine whether time pressure produces parallel effects for speech with and without addressees present.

This language game enables us to examine the interface between message-level planning, or the development of a message to be conveyed, and utterance planning—in this case, the development of referring expressions (Brown-Schmidt & Konopka, 2008; Brown-Schmidt & Tanenhaus, 2006). Smith and Wheeldon (1999, 2001) termed this type of planning, which spans both message-level planning and grammatical encoding, "high-level" planning because it excludes lower-level aspects of planning such as phonological encoding.

Predictions of a Good-Enough Approach

Our good-enough approach to sentence production suggests that production processes are constrained by limited resources. Those resources are allocated in service to particular goals. This perspective allowed us to make predictions for both the goals of speaking promptly and audience design.

Based on prior research (Ferreira & Swets, 2002, 2005), we predicted that the goal to speak quickly should cause speakers to plan less content in advance of speech. To measure planning scope, we determined when speech slowed down in the tangram conditions relative to real-world object displays that excluded tangrams. In the example in Figure 1, a Director did not need to give a description of that tangram until the end of the first clause of the target sentence (e.g., *The shoe moves below the [tangram description]*). In this case, a Director may either expend planning effort in advance, slowing down at the beginning of an utterance to plan the description, or wait until just before articulating the description to plan for the tangram. In either case, we hypothesize that there is a temporal cost to planning descriptions for these tangrams relative to real-world objects. The locus of this cost will help determine the scope of planning.

To this end, we chose two sections of utterances for analysis of planning scope: durations of the Initiation Time and *moves below the* sections seen in Table 1. We chose these sections because they offer us two different ways in which to assess the way speakers were planning for the object being described at position Y1. Whereas initiation time offers information about more advance planning for the Y1 region, the *moves below the* section, immediately preceding the object description, offers information about more incremental planning. We predicted an effect of tangram presence during Initiation Time such that it would take longer to begin speaking if a tangram is in that position rather than a real-world object, indicating that the speaker is planning the description of that object before beginning to speak. However, we predicted that this effect of object type would only arise at initiation time without time pressure. With time pressure, the effect would only be found further down the sentence.

We now turn to the possible effects of the presence of an addressee. At the heart of our project lies the question of how speakers respond to the goals instantiated by speaking to addressees, which would seem to elicit simultaneous goals of speaking quickly and with good detail. Our predictions and analyses begin with the noun section marked Y1 in Table 1, corresponding to the description of the object occupying the center position of the array, such as the tangram in Figure 1. We expect that the presence of an addressee will provide a greater incentive to provide adequate details to an addressee,

TABLE 1
Template for Division of Target Utterances Into Analysis Sections

[Initiation Time] The X [moves below the] [Y1] and the Z moves above the Y2

We show the regions we analyze in brackets. Sections X, Y1, Z, and Y2 represent places at which speakers produced descriptions of either tangrams or real-world objects.

resulting in richer descriptions. Although prior research has found that the presence of addressees who engage speakers in interactive dialogue elicits shorter descriptions (Krauss & Weinheimer, 1966; Murfitt & McAllister, 2001), we did not expect truly interactive dialogue to emerge from this paradigm. Interactive dialogue in general allows speakers to get away with producing shorter descriptions because back-channel feedback indicates to the speaker when the message has been adequately conveyed (Clark & Wilkes-Gibbs, 1986). Without this consistent back-channel, however, the mere presence of an addressee should instantiate a more acutely believed goal to be clear, resulting in longer, more detailed descriptions of tangrams.

If the presence of an addressee only invoked the goal to speak more quickly to avoid awkward, lengthy pauses (Jefferson, 1989), then speakers in the present experiment would show the same types of effects in the addressee present condition as they would in the deadline condition: a tendency to reduce the scope of planning. Although this is an intriguing possibility, we thought it was more likely that a different pattern would emerge. Specifically, addressees bring about dual pressures to speak clearly and quickly, which may somewhat balance each other out. Although speakers with copresent addressees should offer more lengthy descriptions of tangrams, we predict that they will do so without taking additional time to plan these descriptions during the initiation time and *moves below the* sections of the descriptions.

To sum up, the good-enough approach predicts differences in planning strategies based on situational goals. First, the goal of speaking quickly in response to an explicit deadline should cause speakers to plan less of the utterances in advance—to be more incremental. We further expected that the presence of an addressee, who presents a natural time pressure, might produce similar adjustments in the scope of planning but that an accompanying goal to adjust the quality of utterances for this addressee might balance out the two goals of producing utterances for addressees with both speed and sufficient clarity. In short, speakers with addressees should plan more detailed descriptions without sacrificing planning time. Such a pattern of results, in which speakers make more efficient use of limited planning time while addressing a copresent listener, would support the good-enough notion that speakers adjust the dispersal of planning resources depending on the goals at hand.

METHODS

Participants

Participants were Stony Brook University undergraduates who received course credit or $8 for their participation. In the Addressee Present condition, 48 participants served as either Directors or Matchers (24 apiece appeared in each role). In the Addressee Absent condition, an additional 24 participants served as Directors. Because only 15 of the 24 Addressee Absent Directors consented for their voice recordings to be used, we had only 15 Matchers in the Addressee Absent condition.

Materials

Each Director and Matcher engaged in a matching game consisting of 12 rounds. Each of the 12 rounds included two experimental utterances and six filler utterances that enabled Matchers to move objects around in an image grid.

Stimulus creation: Directors. This experiment required visual displays that would prompt participants to produce an utterance in the form of *The X moved above the Y and the Z moved below the Y*, with X, Y, and Z being three different noun phrases used to describe images of objects on a computer monitor. X, Y and Z also correspond to the positions in the arrays, in which X images appeared in the first position, Y in the center, and Z in the third position from left to right. The images were either real-world objects that could be easily lexicalized or tangrams that are not easily lexicalized. We constructed these picture arrays in Adobe Photoshop (Adobe Systems Incorporated, San Jose, CA). We chose 72 real-world objects from a database of color images courtesy of Michael J. Tarr (Carnegie Mellon University, http://www.tarrlab.org) and selected objects that had elicited labels that were one syllable long with high naming accuracy (mean = 94.7%) and low reaction times (mean = 771 ms) as reported in the norming database available in Rossion and Pourtois (2004).

We conducted a norming study to select a group of 24 tangrams, beginning with a set of 48. We asked 11 participants to name the tangrams as they were presented on a computer screen, encouraging them to describe objects with brevity but also with enough detail for someone to identify the object they described within a larger pool of objects. To encourage this unique identification of tangram names (as opposed to names that did not differentiate them from other tangrams, such as "black shape"), the experimenter sat across from the participants and looked for the named objects in a Windows Explorer folder set to "thumbnail view." Participants also named a set of 48 real-world objects from the Rossion and Pourtois database. The experimenter said "okay" when he or she found the tangram or real-world object's thumbnail image. For each tangram we calculated the mean naming time and standard deviation across participants, with naming time defined as the time taken to utter the first word that uniquely described a tangram object. For use in the main experiment, we chose the 24 tangrams with the lowest and least variable naming times. The range of these naming times was 2,393 to 4,999 ms. Note that even the tangrams with the lowest mean initiation times still resulted in longer latencies on average than the real-world objects with the highest initiation times.

As shown in Figure 1, we distributed the real-world objects and tangrams to displays with three items aligned horizontally across the screen. We used each of the 24 tangrams in combination with a unique array of three real-world objects (i.e., the objects in each tangram's array did not overlap with the objects in other experimental arrays). We arranged the real-world objects into display positions based on their reported naming latencies and accuracy (Rossion & Poutois, 2004). The rationale behind this decision was to reduce the variability in initiation times between arrays, so we used the objects with the greatest accuracies and lowest naming latencies (all 100% and < 763 ms) as initial objects in each display, followed by objects with slightly less accuracy and greater naming times in the second position (> 90% and < 970 ms), followed by objects with the lowest accuracy rates in the third position (> 70% and < 921 ms). Across the experiment, each display appeared in one of four versions: with its original three real-world objects or with a tangram replacing the first, second, or third object. We created 96 such displays in all (24 per list; see Design, below).

Figure 1 also exemplifies the arrows that appeared in each display as cues for the Directors' motion descriptions. As suggested by the directions of the arrows, half the displays had the first image moving below the center image and the third above. The other half had the opposite pattern. We also created 288 fillers (72 for each list) to make the Directors' speech planning more diverse and to make the game less predictable for the Matchers. Approximately half the filler displays included the directional commands *next to* (rather than *above* or *below*). In addition, 48 fillers required utterances that were only one clause long (e.g., "The corn moves below the cat").

The content of the fillers and their placement provided extra moves for the matching game. Filler items only contained tangrams if the experimental trial involving that tangram had already appeared. This ensured that the Director did not already have the tangram lexicalized during the experimental trial involving that tangram.

We also created five practice displays using one-syllable, real-world objects that were not present in the experimental displays. These practice trials included examples of all possible movements (above, below, and next to).

We used Ulead (Corel Corporation, Ottawa, ON, Canada) Gif Animator software to create a countdown image for the time pressure condition. The image began with three exclamation points aligned vertically. Based on the results of pilot studies, in which participants' average initiation times were between 2,000 and 2,500 ms, we set the deadline to initiate speech at 2,500 ms. One exclamation point disappeared approximately every 833 ms until the countdown was finished (totaling 2,500 ms).

Stimulus creation: Matchers. For each list of experimental arrays and fillers, we created 12 image grids for each round of the games the Matcher played. Each was a 3 (horizontal) by 6 (vertical) grid created in Powerpoint (Microsoft, Redmond, WA; see Figure 2 for an example). We combined individual images from two sets of item arrays (three objects plus one tangram per set) into each grid. Each grid therefore contained six real-world objects and two tangrams. We intended the presence of two tangrams to motivate Directors to produce descriptions that went beyond such descriptions as “the black shape.” These image grids allowed for the movement of each individual image from one slot to another by clicking and dragging it to an empty slot using a mouse.

Apparatus

Directors used a Dell computer and a Dell LCD monitor (Dell Incorporated, Round Rock, TX) to view the visual displays. DirectRT (Empirisoft Corporation, New York, NY) software presented the images on screen and recorded Directors' speech. Matchers used a Dell laptop to manipulate their game grids, which were presented via PowerPoint.

Design

The experiment used a mixed 2 (deadline presence) × 2 (addressee presence) × 2 (object type) design. We manipulated time pressure (deadline or no deadline) and addressee presence (present or absent) between participants. We manipulated object type within participants. Note that in our analyses, our object type manipulation compared control displays (in which a real-world object rather than a tangram appeared at position Y) to displays in which a tangram appeared in the Y position. We had originally manipulated the presence of tangrams at all three locations of X, Y, and Z. We chose to focus on Y position tangram displays because they required Directors to describe the tangram in the first clause of their instructions (compared with position Z), while also allowing Directors the opportunity for more or less advance planning (compared with position X). That is, Directors could begin planning their description of the tangram before beginning to speak or immediately preceding their description. In short, by comparing the Y-tangram condition with the real-world object condition, our results provide essential tests of our hypotheses in the clearest manner possible.

We grouped each tangram with three real-world objects. From that set of four images, we created four different versions of three-object displays presented across four lists: A tangram-absent real-world object display or displays with tangram in the X, Y, or Z position. Directors saw one version of each set in the list presented to them (i.e., they saw 24 of the 96 resulting displays). Thus, each Director saw six real-world object displays and six Y-tangram displays.

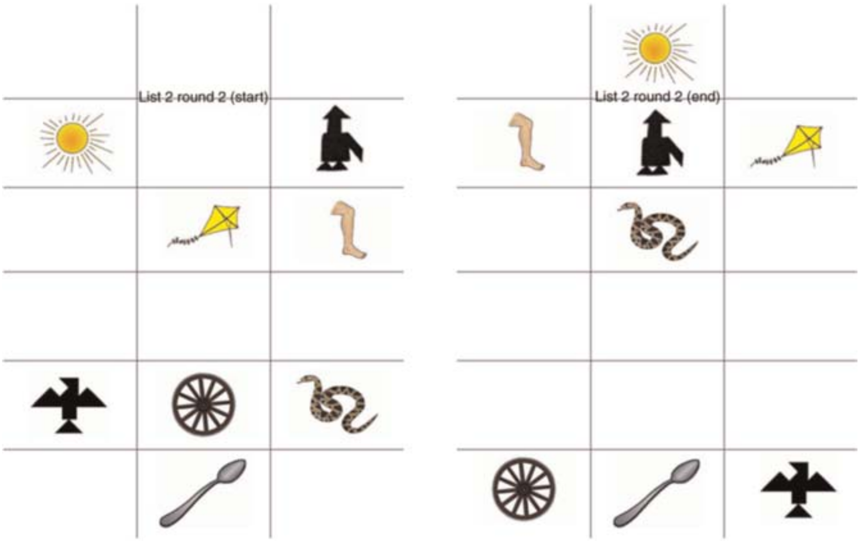

FIGURE 2 Examples of matcher displays. The left side shows the initial state of a display before the beginning of a given round. The right side shows target state of the display after all commands had been given for a round (color figure available online).

Procedure

Addressee present condition. We randomly assigned participants to serve either as Director or Matcher. The participants sat at opposite ends of a large table in front of either a computer monitor (for the Director) or a laptop (for the Matcher). A large foam board partition placed in the middle of the table obstructed participants' views of each other. Directors put on a headset microphone that recorded their speech. After reading some written instructions and having these instructions repeated and clarified by the experimenter, Directors described five practice displays. During this practice session, the experimenter trained the Directors to describe the image arrays to fit the desired utterance format. The experimenter emphasized the importance of accurate descriptions. To clarify to both participants that their views of objects were very different from the other's, the experimenter explained that the Matcher had more objects in his or her view during any one description than the Director did. The experimenter then showed Directors an example of a Matcher grid. The intent of showing the grids to Directors was to make them somewhat aware that simple descriptions of tangrams such as "the black shape" would not suffice for Matchers, given that Matchers might be looking at more than one tangram in their displays.

In the deadline condition we instructed Directors to begin speaking before the exclamation points had disappeared. If the Director did not begin speaking within a 2,500-ms window, a message appeared on the screen after the trial saying *Respond before the countdown finishes!* In the no-deadline condition, we merely verbally encouraged Directors to give their descriptions quickly. This instruction to speak quickly is standard in language production experiments, and previous experiments have shown that the use of this instruction without the presence of an explicit deadline resulted in longer latencies to begin speaking compared with a condition in which the instruction is accompanied by an explicit deadline (Ferreira & Swets, 2002, 2005).

Each trial began with a fixation symbol (a " + " sign) in the center of the screen. When the Director pressed the spacebar, a three-object array appeared that filled most of the screen, centered both vertically and horizontally. The experimenter asked Directors to produce descriptions using the expected sentence frame. Although we allowed Directors to use words such as "over" rather than "above," we used the practice session to discourage larger subversions of the frame, such as descriptions of the images from right-to-left or descriptions beginning with a command such as "Put the shoe below . . . ." The experimenter also told Directors to allow the Matcher time to move the images on their screen before moving to the next trial. Arrays remained on the Director's screen for the duration of a trial. Directors pressed the space bar to indicate they were finished with a trial.

The experimenter showed the Matcher how to move objects around the Powerpoint image grid using the Directors' descriptions as movement commands and explained that we would later compare Matcher boards with target boards. The experimenter also explained to the Matcher that the Director would indicate when the Matcher should move on to the next image grid, beginning the next game. Also, because we were inclined to study the effect of the mere presence of addressees on speech planning, we did not want to encourage a particular kind of interaction between Matcher and Director. For this reason, we did not offer any specific instructions whatsoever about how and whether the Matcher should speak to the Director. That is, we neither encouraged nor discouraged feedback. The result of this lack of instruction was that Matchers only very rarely interacted with Directors.

Addressee absent condition. The experimenter instructed Directors in the Addressee Absent condition to give descriptions such that another participant could later understand the descriptions as commands to move the images on the Matchers' boards. The experimenter did not explicitly tell Addressee Absent Directors that Matchers would eventually be recruited to listen to their descriptions, and the experimenter did not ask these Directors explicitly to give the Matcher time to move their images. In other respects, the instructions for

Addressee Absent Directors were the same as for the Addressee Present conditions.

Addressee Absent Matchers participated in this experiment only after we collected the audio recordings from the Directors and prepared them for use. For Matchers in this condition, the experimenter explained the game the same as in the Addressee Present condition. Instead of listening to a live Director, however, these Matchers heard the voice commands recorded from the Addressee Absent Directors. Matchers listened to the voice commands over headphones. After each command the Matchers attempted to move the correct object or objects in the image grid. Matchers advanced to subsequent voice commands by pressing the spacebar on a keyboard.

RESULTS

Research assistants transcribed all of the Directors' utterances in their entirety. To test our hypotheses we analyzed several sections of each utterance. Table 1 provides the template we use to refer to the parts of the utterances for each data analysis. For each analysis, we used mixed models from the lme4 package (Bates, Maechler, & Bolker, 2012) in R. For analyses of continuous measures such as initiation time, we fitted the data using linear mixed effects models (Baayen, Davidson, & Bates, 2008). For analyses of dichotomous variables, we fitted the data using mixed logit models (Jaeger, 2008). With these approaches, we simultaneously modeled participants and items as random factors. We first present analyses of the Y1 region, which allowed us to draw conclusions about the content of speakers' utterances. We then present analyses of Initiation Time and the *moves above/below the* region, which allowed us to draw conclusions about the time course with which speakers planned that content. Finally, we describe results for Matcher accuracy.

Object Descriptions

Our first analyses ask whether speakers gave longer, richer descriptions of tangrams under different conditions of time pressure and audience presence. To test our predictions, we used word counts and description richness measures. These analyses focus on the noun section marked Y1 in Table 1, corresponding to the description of the object occupying the center position of the array, such as the tangram in Figure 1.

If speakers make strategic adaptations in response to task circumstances, those adaptations should influence the content and duration of descriptions. For tangrams, we predicted that speakers with addressees, due to pressures of audience design, might offer longer and more helpful descriptions than speakers without addressees, whereas time pressure might lead to shorter and less helpful tangram descriptions. For real-world objects, we expected descriptions to remain consistent across all conditions because such objects are always best described with a single, one-syllable word (e.g., shoe).

Object description word counts. We first present word counts for the descriptions of objects offered at section Y1. To generate these tallies, a research assistant who was naïve with respect to the hypotheses of the experiment counted the number of unique words used to describe each object. The assistant excluded determiners, such as *the* and *a* because they did not add any meaningful content to the descriptions. For the same reason the counts also excluded filler-type disfluencies "uh" and "um," which were quite rare. All other types of words were included in the count unless they were repeated. Repeated instances of words were not added to the count of the first instance. For example, the words *it* and *looks* were counted only once in the following utterance: "The truck is going below, it, it looks, it looks almost like a Santa Claus hat and three triangles...." Such repeated instances of words were also quite rare.

We present the data in Figure 3. The linear mixed effects model revealed, to begin, an unsurprising main effect that descriptions of tangrams were longer than those for real-world objects (Table 2). Our more important prediction was that speakers would produce longer descriptions for tangrams in the presence of an addressee. In fact, the analysis revealed an interaction between addressee presence and object type. Directors included more content words in their descriptions of tangrams when Matchers were present, whereas Matcher presence did not affect the length of real-world object descriptions. All other main effects and interactions were not significant. Thus, speakers never adjusted the length of real-world object descriptions and only adjusted the lengths of their tangram descriptions in response to the presence of an addressee.

This analysis indicates that the presence of an audience prompted speakers to produce tangram descriptions that contained more words—irrespective of explicit time pressure. It would be natural to assume that lengthier descriptions were somehow better, but this may not be so. In an additional analysis of the Y1 region, we turned to the particular content of those descriptions.

Description richness. The presence of live addressees in the Addressee Present condition provides circumstances in which speakers should experience more pressure to engage in audience design. One clear way in which speakers could modify their utterances in response to this pressure would be to allude to the fact that a tangram object was from a different category than real-world

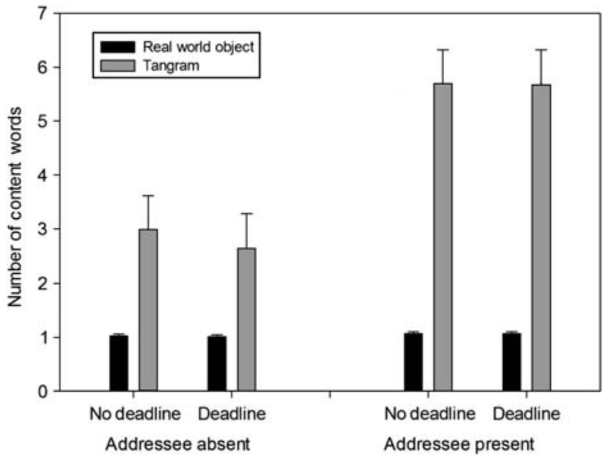

FIGURE 3 Object description word counts as a function of addressee presence, deadline, and object type. Error bars are standard error, $N = 48$.

objects, leading to a description that marked the tangram's presence (e.g., by saying "The thing that looks like a mountain" rather than "a mountain"). The experimental task required Matchers to identify the referents of each description. For that reason descriptions that overtly marked the objects as tangrams were potentially helpful to Matchers because they offered a way of distinguishing the tangram objects from the real-world objects.

Two research assistants who were naïve to the study's hypotheses and blind to the conditions judged whether each Y1 utterance in the tangram condition contained an allusion to a tangram of the sort described in the previous paragraph. There was good interrater reliability between the two coders, $\alpha = .84$, and data from both coders showed the same pattern of effects when subjected to the mixed logit analysis. The data confirmed our expectation that speakers with addressees present were more likely to make such allusions (mean = .76, standard deviation [SD] = .29) than speakers without addressees (mean = .28, SD = .37). Table 3 presents results of the logit mixed model used to analyze results from this dichotomous measure. The analysis confirmed that the main effect of addressee presence was significant and that deadline produced neither a main effect nor an interaction with addressee presence. Taken together, our analyses of the Y1 region suggest that speakers in the presence of addressees were able to plan richer, longer noun descriptions than speakers without addressees present.

TABLE 2
Fixed Effects for the Linear Mixed Model Predicting Continuous Dependent Variables

	Number of Words (Y1)			*Initiation Time*			moves below the *Duration*		
	Estimate	*SE*	*t*	*Estimate*	*SE*	*t*	*Estimate*	*SE*	*t*
(Intercept)	1.07	.46	2.33**	1,651.30	191.70	8.62***	1,141.90	221.71	5.15***
Deadline	.08	.64	.13	−479.40	267.60	− 1.79*	.89	308.01	.00
Addressee	− .04	.64	− .07	− 311.10	267.20	− 1.16	255.25	307.60	.83
Tangram	4.63	.36	12.84***	569.60	132.80	4.29***	1,033.50	198.01	5.22***
Deadline × addressee	− .10	.91	− .11	103.20	377.20	.27	− 352.46	435.64	− .81
Deadline × tangram	.72	.51	1.42	− 620.60	185.50	− 3.35**	− 141.90	280.03	− .51
Addressee × tangram	− 2.66	.51	− 5.22***	− 173.50	184.90	− .94	512.89	279.58	1.84
Deadline × add. × tan.	− 1.04	.72	− 1.45	256.60	260.30	.99	− 252.86	396.08	− .64

SE, standard error.
$*p < .05$, $**p < .01$, $***p < .001$.

Planning Regions

Our first set of analyses focused largely on the content of Directors' descriptions. Our next set of analyses examines the time course with which they planned that content. We focus on two planning regions: Initiation Time and *moves above/below the* (see Table 1).

Initiation Time. Initiation Time represented the time that elapsed from the onset of the image array to the speaker's first vocal response. One research assistant calculated the duration of each section using Praat (Phonetic Sciences, Amsterdam, The Netherlands) software. Figure 4 presents data for this measure, and Table 2 presents the results from the linear mixed-effects analysis. A full table of means in which each cell of the design is represented can be found in the Appendix.

As we expected, Directors began speaking significantly more quickly when they experienced time pressure. Figure 4 also indicates that participants took longer to initiate their utterances when the Y1 object was a tangram rather than a real-world object. The data also reveal a reliable interaction between time pressure and object type. As seen in Figure 4, when speakers experienced a deadline, the initiation times for tangrams and real-world objects were nearly identical. Because the description of the Y1 object constitutes the end of the utterances' first clause, this pattern suggests that the deadline prompted Directors to plan less than a clause in advance. However, without a deadline, Directors' initiation times were longer when tangrams occupied position Y1 than when real-world objects occupied that position. These data suggest that Directors used an increment in Initiation Time to plan about a clause in advance. Thus, when participants were not under time pressure, they seemed to plan to the end of the first clause. However, when time pressure was applied, their planning stopped short of that point. This result suggests that in this paradigm, as in previous paradigms (Ferreira & Swets, 2002, 2005), time pressure influenced the scope of planning.

TABLE 3
Fixed Effects for the Mixed Logit Model Predicting the Binomial Measure of Richness

	Estimate	*SE*	*Wald Z*	*p*
(Intercept)	1.58	.76	2.10	< .05
Addressee	−3.43	1.10	−3.11	< .01
Deadline	.91	1.14	.80	.42
Addressee × deadline	−.59	1.60	−.37	.71

SE, standard error.

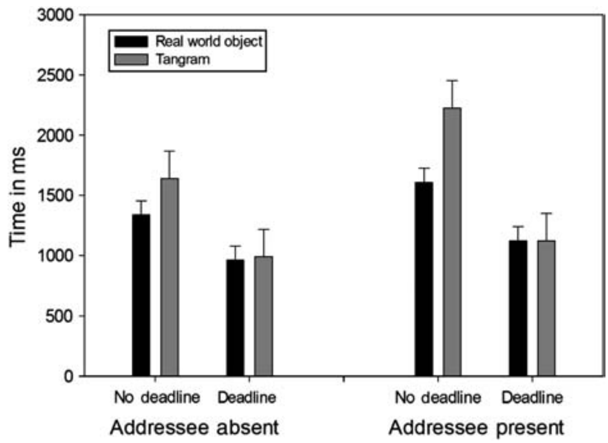

FIGURE 4 Initiation Time as a function of addressee presence, time pressure, and object type. Error bars are standard error, $N = 48$.

Our prior analyses of the content of Directors' descriptions of the Y1 objects provided consistent evidence that speakers in the presence of addressees planned richer, longer noun descriptions than speakers without addressees present. If Directors need more time to plan that richer content, one place at which we might expect to find that additional time would be during Initiation Time. If so, we would expect to obtain Initiation Times in which addressee presence interacted with object type such that initiation times are higher for tangram utterances in the presence of addressees but the same across addressee conditions for real-world object sentences. However, we did not obtain that interaction (Table 2). For another view of this, compare the deadline portions of Figures 3 and 4. In Figure 4, we see that deadlines prompted Directors with or without addressees to initiate their tangram and real-object descriptions with the same time course. Although inspection of Figure 4 might suggest that Initiation Times were generally longer when Directors had addressees present, this effect was not significant. Even so, the Directors produced richer content for the tangram descriptions, especially when addressees were present. Hence, speakers with addressees seem capable of offering more generous tangram descriptions without sacrificing any planning time during the Initiation Time period.

For Initiation Times there were two critical results. The first was that different conditions of time pressure resulted in different degrees of advance planning. The second was that addressee presence did not affect initiation times in patterns that corresponded to the longer descriptions given to tangram objects. Directors did not take longer to begin speaking in the presence of addressee, and certainly did not do so in proportion to the differences that emerged in description lengths.[1]

Moves above/below the. This section comes just after articulation of the noun in Position X and just before articulation of the noun in Position Y1. We suggest that the duration of this utterance section provides an indication of the extent of incremental, phrase-by-phrase planning. Once again, one research assistant calculated the duration of each section using Praat software. We present the data in Figure 5.

We expected that Directors would spend more time articulating this section when the object at position Y1 was a tangram rather than a real-world object. The data support this prediction (see Table 2). This result shows speakers took a lot of time to plan the description when the object that immediately followed was a tangram. It also suggests that speakers engaged in both advance planning (as shown by the Initiation Time effects) and incremental planning (as shown by these *moves above/below* effects) for these utterances. The analysis also revealed a marginal interaction between addressee presence and object type ($p = .073$). No other effects achieved significance.

Disfluencies. In some instances, an analysis of filler-type disfluencies such as *uh* and *uhm* has provided a broader picture of moment-to-moment planning processes than temporal measures alone (e.g., Brown-Schmidt & Tanenhaus, 2006). However, participants in our paradigm were generally quite fluent in their speech: Only 7.2% of participants' utterances contained such disfluencies. When considering disfluencies only at the two planning sections, the frequency of disfluencies dropped to 1.4% during the initiation time window and to 3.4% in the *moves below the* window. Full factorial analyses of disfluencies at these regions using mixed logit models failed to converge. When they did converge, after we removed fixed effect interaction terms and random item terms, the only effect we

[1]As pointed out by an anonymous reviewer, the claim that addressees elicit longer descriptions without sacrificing planning time (or, as we put it here, the claim that effects of addressee presence on initiation time effect do not arise in proportion to the effects of addressee on description length) can be evaluated statistically. We did so by including a covariate term in the linear mixed-effects models for both the description length measure (for which we included a mean-centered initiation time covariate term) and for the initiation time measure (for which we included a mean-centered description length covariate term). Neither model produced significant effects of the covariate terms, and including these terms did not result in significant chi-square differences in the fit of each model. To put it another way, initiation time did not predict description length and its presence as a covariate did not help explain the other effects. It was the same for initiation times: Description length did not predict initiation times, and its presence as a covariate did not explain the other effects.

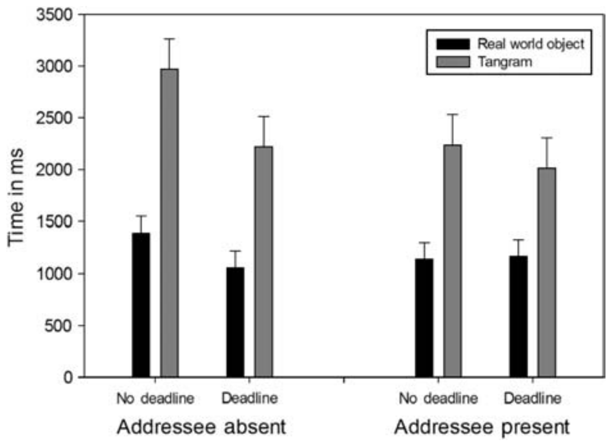

FIGURE 5 *Moves below the* duration as a function of object type, addressee presence, and time pressure. Error bars are standard error, $N = 48$.

found was that participants were more likely to be disfluent in the presence of tangrams. However, we only observed this effect in the *moves below the* analysis. Because the scarcity of disfluencies prevented us from detecting other effects, we focused our analyses of planning regions on temporal measures.

Matcher Accuracy

We assessed Matcher accuracy by coding the proportion of objects placed in their correct positions on a Matcher's board at the end of each round. Recall that there were 12 rounds per Matcher and eight objects per board, which were moved around on eight separate trials during each round. Because we can only assess the accuracy of boards after a given round (but not after a given trial), it is not possible to code the extent to which Matchers successfully moved objects in response to Directors' individual descriptions. Rather, we can only evaluate the overall performance of Matchers after each round. This constraint also renders analyses by individual items (utterances) impossible because each round had eight items. Therefore, we only conducted analyses of variance with participants as the random factor.

There was a tendency for Matchers to place more objects in correct positions in the presence of Directors (mean = .93, SD = .10) than in the absence of Directors (mean = .90, SD = .10). However, unlike previous studies showing advantages of addressees compared with overhearers of commands (Schober & Clark, 1989), the effect was not significant, $F(1, 35) = 1.59$, mean square error (MSE) = .009,

$p = .22$. There was neither an effect of deadline ($F < 1.2$) nor an interaction between deadline and addressee presence ($F < 1$).

DISCUSSION

This study emerged from a good-enough approach to language production, which highlights the impact of speakers' goals on speech planning. We found that the presence of an addressee and explicit time pressure elicited complementary patterns of adjustment. First, the presence of an addressee, with or without explicit time pressure, induced speakers to make utterances more descriptive without significantly changing the time they took to plan their utterances or the scope with which they planned those utterances. The pattern of results from noun descriptions and planning times lead us to conclude that speakers altered the efficiency of their planning under different conditions of addressee presence. Even though speakers took the same amount of time to begin speaking across circumstances of addressee presence, the tangram descriptions offered to copresent Matchers were much richer than those uttered by Directors without copresent Matchers—in number of words and in a measure that coded the helpfulness of the descriptions. In sum, speakers with copresent addressees planned sentences more efficiently than speakers without addressees.

In addition, we found that explicit time pressure, with and without addressees, made speech more incremental in ways similar to results from previous studies (Ferreira & Swets, 2002, 2005). Speakers under different conditions of time pressure showed important differences in the scope of their advance planning. Without time pressure, Directors began planning for the second noun phrase before they began to speak. Under time pressure, Directors did not extend their planning as far in advance. The finding that speakers thus adopted a longer scope of high-level planning when they felt no explicit time pressure is consistent with flexibly incremental accounts of language production (Ferreira & Swets, 2002, 2005) and adds to such previous findings by showing that these effects are robust to the goals elicited by the presence of addressees. In the remainder of this discussion, we consider how our results might help establish some early principles of a good-enough perspective on language production.

Good-Enough Language Production

We propose that the language production system, like the language comprehension system, may deploy planning resources in a "good-enough" manner. Our proposal draws on several of the insights and themes raised by the good-enough approach to language comprehension. Essentially, we claim that, like sentence comprehension, sentence production is resource constrained and can adjust the dispersal of these limited resources in service to particular goals. There are two themes of the good-enough approach that we address in turn. The first concerns the limited availability of resources and whether evidence exists that planning processes might be constrained by such limitations. The second theme concerns the idea that these limited resources can be deployed in goal-oriented, flexible ways when faced with the real pressures people experience in circumstances of dialogue.

Resource limitations. One of the central motivations for the good-enough approach to sentence comprehension was the idea that the parsing mechanism does not have unfettered access to the mental resources required for addressees to comprehend sentences fully (Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira & Patson, 2007). Rather, the limited bandwidth available for understanding a sentence imposes constraints on the extent to which possible sources of information might be consulted and the depth with which various possible parses of ambiguous input might be considered. Although there is a rich research tradition investigating the ways in which memory resources constrain sentence comprehension, investigations into the relationship between cognitive resources (primarily working memory) and sentence planning have not been as common. Several studies have demonstrated that higher-level language production planning is constrained by working memory (Hartsuiker & Barkhuysen, 2006; Horton & Spieler, 2007; Kellogg, Oliver, & Piolat, 2007; Kemper, Herman, & Lian, 2003; Kemper & Sumner, 2001; Mortensen, Meyer, & Humphreys, 2008; Schriefers & Teruel, 1999; Slevc, 2007; Wagner et al., 2010). The most direct evidence for the impact that working memory limitations has on planning comes from other work from our laboratory (Swets, Jacovina, & Gerrig, 2012). In that project, individuals with higher working memory capacity were more likely to set a longer scope of planning than individuals with a lower working memory capacity. The mounting evidence from such studies indicates that working memory is involved in language production to a nontrivial degree, suggesting that the quality and scope of speech planning may well be subject to the kinds of resource constraints assumed in the good-enough perspective.

Responsiveness to goals elicited in dialogical circumstances. Another primary motivation of the good-enough approach to comprehension is the idea that language does not take place in the idealized circumstances provided in typical laboratory settings. Rather, language typically operates in dialogical circumstances in which comprehension and production take place simultaneously and compete with each other for the limited cognitive resources summarized just above. In such circumstances the goals of comprehension shift from delivering a veridical parse to comprehending enough of an interlocutor's turn to generate one's own response

(Ferreira et al., 2002). We designed the present experiment to examine the impact that a copresent addressee has on speech planning. We reasoned that speakers with addressees should neither have the luxury of regularly offering poorly planned, ambiguous descriptions nor of taking as long as they want to begin speaking. Rather, they have simultaneous goals of planning carefully and speaking quickly, not to mention the challenge of preparing much of their own conversational turn while also trying to comprehend the sentences of the interlocutor (Ferreira et al., 2002).

The current experiment demonstrated that when the goals of rapid and careful planning are manipulated concurrently by the presence of an addressee, a speaker will provide more detailed descriptions without sacrificing a proportional amount of planning time. There is an element of efficiency to be observed here, but of a different type from that observed by Ferreira and Swets (2002). Whereas Ferreira and Swets (2002) observed increased efficiency by seeing the same content planned with less time, we present a case in which more content is planned in roughly the same time, regardless of the presence of explicit time pressure. We argue that these results reflect the dual pressures brought about by the attempt to accommodate a copresent addressee, who requires information to arrive both promptly and with clarity. Such results are in line with proposals suggesting that planning sentences seems "easier" in conversation (Garrod & Pickering, 2004; Pickering & Garrod, 2004). However, whereas these previous accounts have mostly attributed the ease of planning in conversation to shared representations and priming between two people working in tandem, we propose that conversational goals elicited by the mere presence of an addressee bring about an incentive for the speech production system to use its available resources with maximum efficiency. Rather than plan only carefully, or only quickly, speakers in conversation rise to accomplish both goals at once.

Future Directions

Many factors make up the difference between circumstances of producing sentences in monologue versus dialogue. This project only investigated issues related to time pressure and the presence of addressees, specifically whether the mere presence of addressees might engender the same goals that time pressure does. The circumstances in the present study, although moving toward a more dialogue-like situation than typical experiments, fell short of the full range of factors applied during dialogue (Pickering & Garrod, 2004). Matchers surprised us by rarely contributing feedback to our Directors. The presence of feedback, such that the Matchers had said, for example, *uh-huh* when speakers had provided a sufficient tangram description, might have altered speech content and planning in ways we were not able to measure (Clark & Wilkes-Gibbs, 1986). Because it allows speakers to include fewer details once enough information has been provided, such feedback might have caused tangram descriptions to be more equal between conditions, if not shorter under more dialogic conditions. In fact, previous studies of differences in the quality of descriptions produced under conditions of dialogue versus monologue revealed that speakers in dialogue tend to produce shorter descriptions than speakers in monologue (Murfitt & McAllister, 2001). Still, the specific question we addressed was not one of the role of feedback but of the effect that a copresent addressee would have on the quality of the content and incrementality of speech planning. Our prediction that speakers would give clearer, longer descriptions with a copresent addressee was confirmed.

Because we did not elicit feedback from addressees, we can only draw limited conclusions about the manner in which dialogical factors influence speech planning. Researchers interested in pursuing how the good-enough approach maps onto sentence planning in dialogue may wish to explore more of the goals that differ between monologue and dialogue. For example, in addition to the role of feedback, future research might also investigate how the properties of turn-taking in ordinary conversation might alter the way sentences are planned (Sacks et al., 1974). Such research might find that genuine competition for the floor in the context of a more natural conversation might affect incremental sentence planning as experimentally induced time pressure does. In such a case it is possible that the dual goals of speaking quickly and planning detailed descriptions might manifest themselves by eliciting speech that is both more efficient *and* longer in scope. In other words, perhaps moving toward more and more natural conversational contexts saves the language production system from the resource costs of time and mental bandwidth simultaneously. In that case, speech might have a longer scope of planning, without costs in time, and with maximum benefit to the quality of content.

Conclusions

In sum, we propose that the language production system, like the language comprehension system, may operate in a manner such that planning resources are flexibly distributed to accomplish particular goals. When there is little incentive for a speaker to provide detail to an addressee, such details are omitted. When a copresent addressee needs specific details, a speaker will provide more detailed descriptions without sacrificing planning time. This pattern of results reflects a system that only deploys just enough resources in the most cost-effective amount of time to get a particular job done. In circumstances that approach the natural context of conversation, the language production system may be pushed to use those resources more efficiently than in circumstances that veer further away from conversation. We argue that this set of results may help point the way to considering language production as a good-enough process (Ferreira et al., 2002; Ferreira & Patson, 2007).

These results have broader implications for models of language production. We suggest, in line with a growing number of researchers, that it is most fruitful for researchers to consider the types of circumstances that change the way in which sentence planning unfolds over time (Christianson & Ferreira, 2005; Costa & Caramazza, 2002; Damian & Dumay, 2007; Ferreira & Swets, 2002, 2005; Korvorst et al., 2006; Schriefers & Teruel, 1999; Wagner et al., 2010). Continuing such research will offer a clearer picture of the workings of a decidedly flexible, dynamic system.

ACKNOWLEDGMENTS

This material is based on work supported by National Science Foundation Grant 0325188. We thank Bill Wenzel for his significant contributions to the collection and analysis of data. We also thank Susan Brennan, Arthur Samuel, and two anonymous reviewers for their helpful comments during the development of this project.

REFERENCES

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. Retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.

Brown-Schmidt, S., & Konopka, A. E. (1998). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, *109*, 274–280.

Brown-Schmidt, S., & Tanenhaus, M. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, *54*, 592–609.

Christianson, K., & Ferreira, F. (2005). Planning in sentence production: Evidence from a free word-order language (Odawa). *Cognition*, *98*, 105–135.

Christianson, K., Williams, C., Zacks, R., & Ferreira, F. (2006). Younger and older adults' good enough interpretations of garden-path sentences. *Discourse Processes*, *42*, 205–238.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: APA Books.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, *58*, 332–373.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J.-F. Le Ny & W. Kintsch (Eds.), *Language and comprehension* (pp. 287–299). Amsterdam, The Netherlands: North-Holland.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1–39.

Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: Implications for the scope of phonological encoding in speech production. *Journal of Memory and Language*, *46*, 178–198.

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, *57*, 195–209.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11–15.

Ferreira, F., & Patson, N. (2007). The good enough approach to language comprehension. *Language and Linguistics Compass*, *1*, 71–83.

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 & Language*, *46*, 57–84.

Ferreira, F., & Swets, B. (2005). The production and comprehension of resumptive pronouns in relative clause 'island' contexts. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 263–278). Hillsdale, NJ: Lawrence Erlbaum Associates.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*, 8–11.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*, B1–B14.

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

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, *47*, 589–606.

Horton, W. S., & Spieler, D. H. (2007). Age-related differences in communication and audience design. *Psychology and Aging*, *22*, 281–290.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), *Conversation: An interdisciplinary perspective* (pp. 166–196). Clevedon, UK: Multilingual Matters.

Kellogg, R. T., Oliver, T., & Piolat, A. (2007). Verbal, visual, and spatial working memory in written language production. *Acta Psychologica*, *124*, 382–397.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201–258.

Kemper, S., Herman, R. E., & Lian, C. H. T. (2003). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech or noise. *Psychology and Aging*, *18*, 181–192.

Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging*, *16*, 312–322.

Kingsbury, P., Strassel, S., McLemore, C., & McIntyre, R. (1997). *CALLHOME American English transcripts, LDC97T14*. Philadelphia, PA: Linguistic Data Consortium.

Korvorst, M., Roelofs, A., & Levelt, W. J. M. (2006). Incrementality in naming and reading complex numerals: Evidence from eyetracking. *Quarterly Journal of Experimental Psychology*, *59*, 296–311.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*, 343–346.

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

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*, B25–B33.

Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2008). Speech planning during multiple-object naming: Effects of aging. *Quarterly Journal of Experimental Psychology*, *61*, 1217–1238.

Murfitt, T., & McAllister, J. (2001). The effect of production variables in monolog and dialog on comprehension by novel listeners. *Language and Speech*, *44*, 325–350.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–225.

Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language*, *30*, 385–406.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*, 217–236.

Russell, A. W., & Schober, M. F. (1999). How beliefs about a partner's goals affect referring in goal-discrepant conversations. *Discourse Processes*, *27*, 1–33.

Sacks, H., Schegloff, E. A., & Jefferson, G. A. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, *50*, 696–735.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Science*, *6*, 382–386.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*, 211–232.

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

Slevc, R. (2007). *Saying what's on your mind: Working memory effects on syntactic production*. Doctoral dissertation, University of California, San Diego, CA.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, *73*, 205–246.

Smith, M., & Wheeldon, L. (2001). Syntactic priming in spoken sentence production—an online study. *Cognition*, *78*, 123–164.

Swets, B., Desmet, T., Clifton, C., Jr., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*, 201–216.

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2012). Individual differences in the scope of speech planning: Evidence from eye movements. (Manuscript submitted for publication)

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 Experimental Psychology*, *36*, 423–440.

APPENDIX

Planning Region Means (in Ms) and Standard Deviations (in Parentheses)

		Control	*Y-Tangram*
Initiation Time			
Addressee absent	No deadline	1,345 (353)	1,620 (607)
	Deadline	974 (229)	991 (339)
Addressee present	No deadline	1,658 (717)	2,113 (1,124)
	Deadline	1,156 (275)	1,132 (276)
Moves below the			
Addressee absent	No deadline	1,390 (866)	2,969 (1,393)
	Deadline	1,056 (405)	2,219 (1,159)
Addressee present	No deadline	1,139 (344)	2,238 (616)
	Deadline	1,166 (471)	2,015 (656)