Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 54 (2006) 592–609
Journal of Memory and Language
www.elsevier.com/locate/jml

Watching the eyes when talking about size: An investigation of message formulation and utterance planning ☆

Sarah Brown-Schmidt *, Michael K. Tanenhaus

Department of Brain and Cognitive Sciences, University of Rochester, New York, USA

Received 29 May 2005; revision received 15 December 2005
Available online 24 February 2006

Abstract

In two experiments, naïve participants took turns telling each other to click on a target picture while gaze was monitored. Critical trials included a *contrast* picture that differed from the target only in size. In both experiments, the timing of speakers' fixations on the contrast predicted whether the contrast was encoded in a phrase with a pre-nominal adjective (*the small triangle*) or a post-noun repair (*the triangle. . .small one*). In Experiment 1, fixations to the contrast were delayed for adjectives in post-nominal phrases (*the square with small triangles*). In Experiment 2, which used a more complex display, delayed gaze at the contrast was correlated with use of a pre-nominal modifier in disfluent productions (*thee uh small horse*). The results provide insight into the interface between message formulation and utterance planning. They also support the hypothesis that one role of disfluency is to provide speakers with time to reformulate messages and utterances.

Keywords: Language production; Conversation; Eye-tracking; Message formulation; Adjective; Scalar

One of the most important developments in language processing during the last two decades has been the emergence of detailed models of utterance planning. These models seek to explain the growing body of evidence about how speakers retrieve lexical concepts, build syntactic structures, and translate these structures into linguistic forms (Bock, 1995; Dell, 1986; Indefrey & Levelt, 2004; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). However, much less is known about how speakers plan and update the non-linguistic thoughts, or *messages*, that are translated into utterances during language production. And, little is known about how message formulation and utterance planning are coordinated. Filling this lacuna in the literature takes on increased importance as psycholinguists begin to extend investigations of real-time processing to interactive conversation (cf. Brown-Schmidt, Campana, & Tanenhaus, 2005; Pickering & Garrod, 2004; also see the papers in Trueswell & Tanenhaus, 2005). In this most basic arena of language use (Clark, 1991), speakers often update messages on the fly based on new insights, new information, and feedback from addressees, all of which can be concurrent with the speaker's planning and production of

☆ This research was partially supported by NIH Grant HD 27206 to M.K. Tanenhaus. Thanks to Dana Subik, Carol Faden, Courtney Pooler, and William Palin for assistance with data collection and analysis.

* Corresponding author. Present address: Department of Psychology, 2155 Beckman Institute, 405 N. Mathews Ave., University of Illinois, Urbana, IL 61801, USA.
E-mail address: brownsch@uiuc.edu (S. Brown-Schmidt).

0749-596X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2005.12.008

utterances. Thus message formulation and utterance planning are interwoven in time and must communicate with one another at a relatively fine temporal grain.

While the interactivity of natural conversation suggests that we plan messages and formulate utterances in a highly incremental fashion, slips of the tongue, and planned speech (e.g., jokes, memorized sentences), demonstrate that planning of an entire utterance or phrase can sometimes be initiated before speaking. For example, exchange errors such as *fl*eaky *squ*oor (squeaky floor, Meyer, 1992) suggest that in these cases, the (phonological) representations of the two words are concurrently active (Dell, 1986; Dell & O'Seaghdha, 1992). Additionally, experimental results from paradigms that encourage speakers to plan full utterances before speaking (for example, Experiment 1 of Ferreira & Swets, 2002; Sternberg, Knoll, Monsell, & Wright, 1988) indicate that speakers can pre-plan some of what they intend to say.

Issues about incrementality and planning in speech production are not limited to modern academic discourse. Wilhelm Wundt and Hermann Paul debated their views on the production of sentences for almost 40 years (for a lively account of this exchange, see Blumenthal, 1970). Paul's (1880) theory of production was influenced by his view that language comprehension proceeds incrementally. He proposed that for speaking, the mental constructs underlying the words of a sentence are prepared sequentially, isomorphic to spoken forms. In contrast, Wundt (1900) viewed the process of sentence production as originating from a wholistic conceptualization, which is transformed into a sequence of words, resulting in a process both simultaneous and sequential (Blumenthal, 1970).

More recent work on incrementality and language production focuses on planning at syntactic, semantic, and phonological levels. Here, we will limit our discussion to the production of complex phrases, which is the focus of the present research. Levelt and Maassen (1981) asked Dutch-speaking participants in their third experiment to describe scenes with moving shapes. When the shapes moved in the same direction, speakers were faster to begin speaking when they used the Dutch equivalent of a phrase like *the triangle goes up and the circle goes up* than when they used a complex noun phrase like *the triangle and the circle go up*, despite the fact that the complex noun phrase was the more frequently used form. This latency difference was taken as evidence that the unit of advance planning includes each of the lemmas in the subject noun phrase. In related work, Ferreira (1991) asked speakers to produce sentences with complex subject or object noun phrases using a memorization task. She found that speech latencies were affected by complexity of subject but not object noun phrases and took this as evidence that the phonological representation of the subject noun phrase is prepared before speaking (but see discussion in Meyer, 1996). Similarly, Costa and Caramazza (2002) found that naming latencies were facilitated by distractors that are phonologically related to the head noun in constructions such as *the car* and *the red car*, suggesting that the phonological representation of the head noun is activated before speaking, regardless of whether it is the first or second phonological word.

However, other findings suggest that only the first word is prepared before speech onset. For example, Meyer (1996) asked Dutch speakers to perform a picture naming task with auditory distractors. Naming latencies were delayed by a distractor that was phonologically related to the first noun in a noun phrase, but not the second (e.g., the Dutch equivalent of *the arrow and the bag*). Distractors semantically related to either the first or second noun decreased latencies. In a more extreme case, Schriefers and Teruel (1999) presented phonological distractors as German-speaking participants named simple pictures such as a red table (e.g., *roter tisch*). Phonological distractors related to the first syllable of the first word facilitated naming times; no effect was found for distractors related to the second word, and distractors related to the second syllable of the first word only had a weak facilitation effect. These results were interpreted as evidence that in some cases, articulation may begin before the first phonological word is entirely planned. Taken together, these results suggest that the degree to which speakers pre-plan a complex phrase is highly variable. One explanation for this variability is that the amount of advance planning varies depending on the task and type of utterance (Schriefers & Teruel, 1999). Another explanation is that planning is partially under the control of the speaker (Ferreira & Swets, 2002).

The work on planning of complex phrases has clearly led to advances in our understanding of how syntactic and form-based planning are executed. However, little work has addressed how messages are prepared. One reason why our understanding of message formulation lags behind that of utterance generation, is that it is difficult to study message formulation experimentally. Collecting empirical observations about message formulation, and how it interfaces with utterance planning, requires creating conditions in which a speaker's message is constructed as she speaks. This constraint rules out experimental paradigms in which the content of the message is tightly controlled or does not need to be updated after the onset of the utterance, as in reading aloud, producing memorized sentences, and naming objects or, in some circumstances, even describing scenes.

A promising line of inquiry comes from research that adapts the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to language production (e.g., Bock, Irwin, Davidson, & Levelt, 2003; Griffin & Bock, 2000; also see chapters in

Henderson & Ferreira, 2004; Levelt et al., 1999). In a typical study, participants are presented with a simple scene rendered as a line drawing. Eye movements are monitored as the participant apprehends the display and plans and produces an utterance describing the scene. For example, Griffin and Bock (2000) asked participants to describe depictions of simple events, such as a woman shooting a man, or lightning striking a church. The sequence of eye movements reflected the order of constituents in the utterance. Speakers looked at pictured objects about 800 ms to 1 s before naming them. This eye-voice lag is similar to the time it takes to initiate naming an object in isolation (Rossion & Pourtois, 2004; Snodgrass & Yuditsky, 1996), suggesting that the eye-voice delay reflects word preparation. The delay between gaze and speech is similar when participants name an array of objects (Griffin, 2001; Levelt et al., 1999). Additional support for the link between gaze and speech comes from work by Meyer and colleagues using a dual object naming task (Meyer, Sleiderink, & Levelt, 1998; also see van der Meulen, 2001) in which they found that viewing times increased for objects with low-frequency names, implicating a link between retrieval of the phonological form of an object name, and gaze.

The pattern and timing of eye movements as an utterance unfolds could, in principle, provide insights into the interplay between message formulation and utterance planning. In practice, however, eye movements have been relatively uninformative about message formulation. One reason is that for simple scenes, the information for planning a message can be apprehended within a few hundred milliseconds (Bock et al., 2003), well within the duration of a single fixation. Thus, work investigating the gaze–speech link may typically conflate message formulation with utterance planning. In the experiments presented by Bock and colleagues (2003), Dutch and English speakers were asked to say the time pictured on a digital or analog clock display. In their second experiment, speakers saw the clock display for either 100 or 3000 ms. Surprisingly, time-telling errors did not rise substantially with shorter viewing times, suggesting that 100 ms was enough time to apprehend the relevant information from the scene. This finding also suggests that the relationship between gaze and speaking may play a facilitatory rather than a necessary role in picture naming and scene description (see Griffin, 2004b). However, scene-related eye movements may have continued after the scene was removed in the 100 ms condition (e.g., Richardson & Spivey, 2000). The two conclusions from the Bock et al. (2003) work that are most relevant to the present research are that at least for simple scenes, apprehension of the scene is fast compared to the relatively slower process of utterance formation, and that the observed gaze-speech link may reflect both message-formulation and utterance planning processes.

One way to disentangle message formulation from utterance planning is to create situations in which the speaker, while in the process of planning or producing an utterance, encounters new information that requires revising the message. If eye movements can be used to infer when the speaker first encounters that information, then the timing between the uptake of the new information and the form of the utterance might shed light on the interface between message planning and utterance planning. In the experiments reported here, we exploited the properties of scalar adjectives to create these conditions.

Speakers will typically use a scalar adjective, such as *big* or *small*, only when the relevant referential domain contains both the intended referent and an object of the same semantic type that differs along the scale referred to by the adjective (Sedivy, 2001, 2003; Gregory, Joshi, Grodner, & Sedivy, 2003). The presence of a scalar-contrast item in the referential domain encourages elaboration of the referring expression in order for the intended referent to be identified (Olson, 1970; Osgood, 1971). For example, in a display of animals with only one horse, a large horse would be referred to as *the horse*. If, however, the display also contained a smaller, but otherwise identical horse, that same picture would be referred to as *the large horse*.

In our experiments, two partners were each seated in front of a computer display with the same set of 12 or more pictures. The target picture was highlighted on the speaker's display but not the addressee's display. Partners played a simple language game, taking turns telling each other which picture was the target. No restrictions were placed on the form of the participant's utterances. On some trials, the display included a contrast picture that differed from the target only in size (e.g., a small horse paired with a large horse, or a small triangle paired with a large triangle).

We begin by comparing production of the names of *simple shapes* that naturally lend themselves to a description with a pre-nominal adjective (e.g., *the large triangle*), and *complex shapes* that are most naturally described using a post-nominal prepositional phrase (e.g., *the triangle with small diamonds*). We evaluated three hypotheses, each of which makes progressively stronger claims about the relationship between fixations to the contrast and the form of the utterance. The first hypothesis is that use of a size adjective will depend upon whether or not the speaker has made a saccadic eye movement to fixate on the contrast picture. The highlighted target should initially attract the speaker's attention, and the number of items in the display exceeds the limits of visual working memory (Vogel, Woodman, & Luck, 2001). Thus, the first fixation to the contrast should provide an estimate of when the speaker first encounters information that size must be included in the message. This hypothesis can be evaluated by exam-

ining fixations to the contrast on trials when the speaker produced a fluent referring expression. If the first fixation to the contrast provides an estimate of when the contrast is first encoded, modification rates should be low when the contrast is present but the speaker does not look at it.

If modification is closely tied to fixations, we can evaluate a second, stronger hypothesis, which is that there will be a systematic relationship between the *timing* of the first look to the contrast and the form of the utterance. For example, when describing a simple shape when a size-contrast is in the scene, speakers might only be able to produce an utterance with a pre-nominal adjective if they looked at the contrast some time prior to the onset of the utterance. When the speaker notices the contrast after having planned a message that does not include size, the message must be identified and repaired and the utterance plan modified. This might result in a post-noun repair to add modification, or perhaps a disfluency to buy time to incorporate the adjective into the noun phrase. Thus, the relationship between the *form* of the referring expression, and the timing between the first fixation to the size contrast and the onset of the referring expression could provide a window into the interplay between message formulation and utterance planning.

A systematic relationship between the timing of the first fixation to the contrast and the form of the utterance would make it possible to evaluate a third hypothesis about the size of the message that is passed on to utterance planning, and the timing between repairs to the message and utterance planning. For example, let's make the simplifying assumption that messages can be updated continuously when relevant new information is encountered. Let's further assume that a message that contains information for a complete referential description is mapped onto an utterance plan before the utterance begins. Then for both simple shapes and complex shapes, the timing between the first fixation to the contrast and the onset of the utterance would be similar. Speakers might, however, pass messages onto utterance planning in smaller units, perhaps because this would facilitate incremental preparation of components of messages as they become relevant to utterance planning. If so, speakers might delay preparation of size adjectives for complex shapes because size adjectives tend to occur in a post-nominal prepositional phrase. This would be reflected in shorter delays between the first fixation to the size contrast, and utterance onset for complex shapes compared to simple shapes.

Experiment 1

This experiment examined the planning and production of modified referring expressions such as *the small triangle* and *the square with small triangles*. We used displays that sometimes included contrasting pairs of simple and complex shapes. For a simple shape, such as a triangle, the contrast was between shapes that could be described in a noun phrase with a pre-nominal adjective, such as *the small triangle* or *the large triangle*. For a complex shape, such as a square with embedded triangles, the contrast was between the embedded shapes, requiring a noun phrase with a post-nominal prepositional phrase, such as *the square with small triangles* or *the square with large triangles*. Importantly, the size modifier, if used, occurs later in the noun phrase for complex shapes than for simple shapes. This experiment allows us to investigate the three hypotheses previously outlined in the introduction: first, fixations to the contrast should predict whether or not the speaker uses a size modifier, with higher rates of modification when the contrast has been fixated. Second, the timing of the first fixation to the contrast with respect to the onset of the utterance should predict whether the speaker uses a pre-nominal adjective or an adjective in a post-noun repair phrase, with shorter lags associated with pre-nominal modification. Third, the timing of looks to the contrast and use of a pre-nominal versus a post-noun repair for descriptions of complex and simple shapes should provide insight into the size of message planning units, and how repairs to these units affect utterance planning.

Method

Participants

Eighteen pairs of participants who were members of the undergraduate community at the University of Rochester were paid for their participation. Each participant was a native speaker of North American English, and all pairs identified themselves as friends.

Procedure

Participants were each seated at a separate computer, separated by approximately five feet. The equipment was arranged so that the participants could not readily make eye contact. One participant wore a lightweight, head-mounted ASL brand eye-tracker, and both participants wore headset microphones. An audio record of both participants' voices, as well as the video-record from the scene camera with gaze position superimposed were recorded to frame-accurate digital videotape at 30 Hz.

A different visual display was presented on each of 288 trials. On each trial, the displays on the two screens contained the same pictures. However, pictures appeared in different positions on the screens to discourage the participants from using a coordinate system to describe the target (e.g., *click on the second shape from the top on the right-hand side*). An example of a speaker and listener display for one trial is illustrated in Fig. 1A

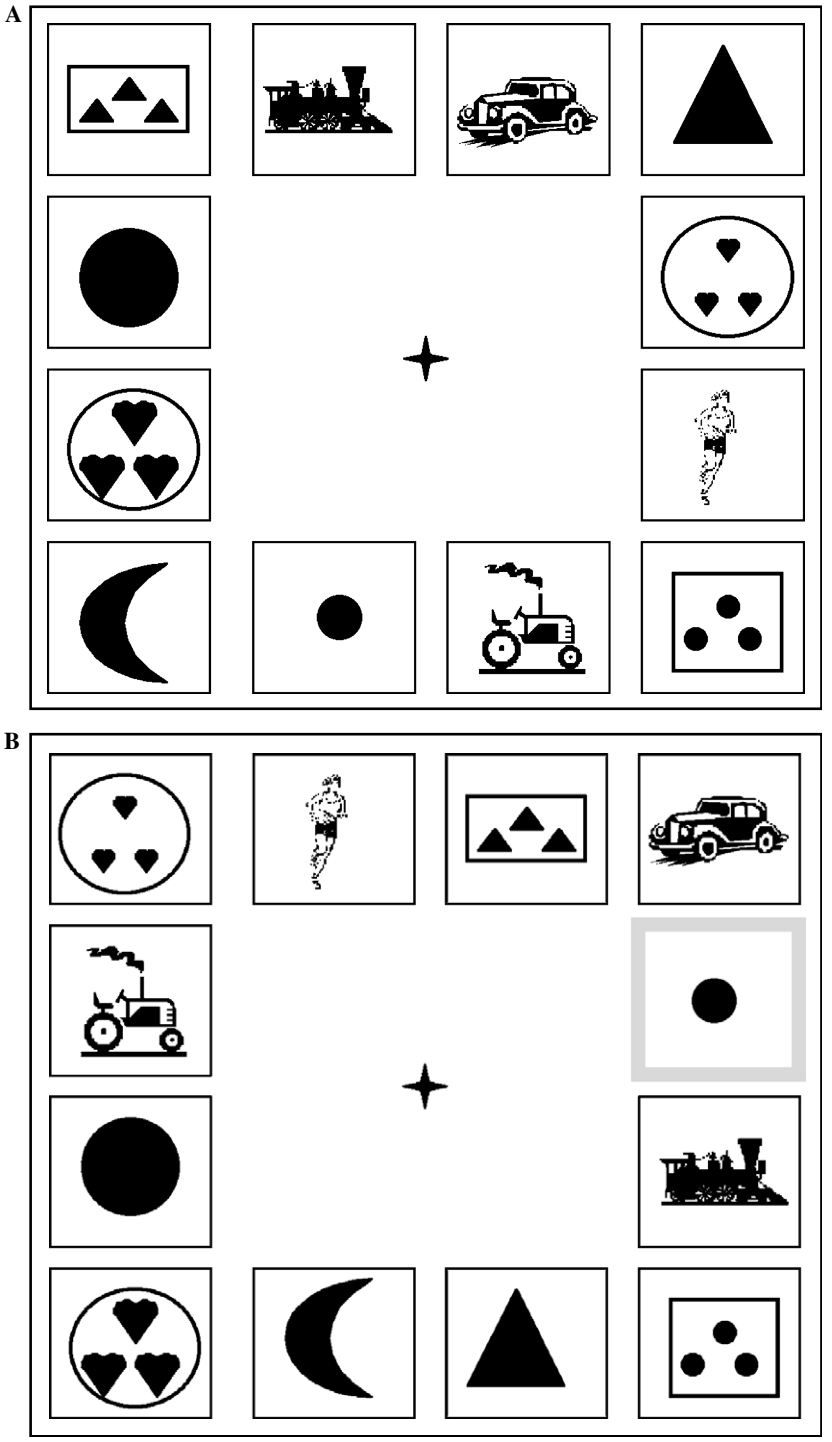

Fig. 1. (A and B) Example stimulus display from Experiment 1. (A) The listener's perspective, and (B) the speaker's perspective.

and B. Participants took turns instructing one another to click on one of the pictures on the screen. One of the pictures was highlighted in yellow on the speaker's screen to designate it as the target. Participants were instructed to tell their partner to click on this item. The experiment lasted approximately one hour.

Two types of experimental target items were used. The simple shapes were objects such as triangles and squares.

The complex shapes had shapes embedded within them, for example, a square with triangles. Simple shapes were used to elicit simple noun phrases e.g., *the triangle*, and *the small triangle*. Complex shapes were used to elicit noun phrases with prepositional phase modifiers that could contain a pre-nominal modifier, e.g., *the square with triangles* or *the square with small triangles*. The shapes were intermixed with pictures of objects such as a camel, rooster, bell, and hammer. A complete list of complex and simple shapes is provided in the Appendix.

Each screen contained twelve pictures: four simple shapes, four complex shapes and four objects. The location of the pictures on the screen was randomized separately for speakers and listeners. On 144 of the 288 trials, the target picture was a simple object. These trials were included to distract participants from the size manipulation. Of the remaining 144 trials, 72 targets were simple shapes, and 72 targets were complex shapes.

On half of the shape trials, a size contrast shape was present in the scene, such as a large triangle if the target was a small triangle or a square with large triangles if the target was a square with small triangles. Thus, on 25% of trials, the target was presented with a size contrast. On the speaker's screen, the distance between the target and the size contrast was manipulated to increase the likelihood that there would be some variation in when speakers would first notice the size contrast. For trials where a size contrast was present, either one or five pictures separated the target and size contrast.

When a complex shape was the target, the scenes included a picture that was a different shape, but contained the same type of small shape, e.g., a circle with small triangles when the target was a square with small triangles. This was done to discourage the use of simple noun phrases for the complex shapes such as *the small triangles*. A shape-contrast was not necessary for simple shapes because the presence of the three other shapes automatically necessitated use of a shape term. Each shape could appear in four different sizes (extra small, small, medium, and large). Thus, speakers had to use the shape's relative size rather than absolute size when describing the shapes.

The 288 trials were randomly ordered into a single list order. The nine exemplars of simple shapes and the nine exemplars of complex shapes were rotated through three different size contrast conditions (e.g., extra-small vs. small, small vs. medium, and medium vs. large) and the two contrast-distance conditions (far vs. close), resulting in six lists. Each pair of participants was presented with a single list. The eye-tracked and non eye-tracked participants each described an equal number of target referents across the different conditions.

Results

A post-experimental questionnaire indicated that most participants thought the experiment was about visual search for the simple objects or about the eye-tracked partner's interpretation of her partner's utterances.

For our analyses, we selected trials for which the eye-tracked participant was the speaker, and the target referent was either a simple or a complex shape. We excluded any trials on which the speaker did not respond or there was equipment malfunction ($n = 17$), leaving 1279 trials for further analysis. In the inferential statistics reported below, our analyses are limited to the participants and items for which we collected data in each of the conditions of comparison. Participants or items were sometimes not included in an analysis if, for example, a speaker did not produce an utterance in one of the conditions while the eye-tracker was functioning properly.

Referential form

An analysis of the referential forms generated by the speakers showed that our manipulations were successful in eliciting the expected forms. Size adjectives were typically used only when a contrast was present in the display. Pre-nominal adjectives were used for simple shapes and adjectives in a post-nominal prepositional phrase were used for complex shapes.

When the display contained a size contrast for the target, speakers used a size adjective on 98% of the trials compared to 27% for trials without a contrast. The difference in modification rates for trials with and without size contrasts was significant (see Table 1 for all F values and related statistics for Experiment 1), however the difference in modification rate between simple and complex shapes did not approach significance. The interaction between presence of a size contrast and shape type was marginal in the participants analysis. For trials with a contrast, there was only a 2% difference in modification rate for simple (99%) and complex shapes (97%). In contrast, for trials without a contrast, speakers were 8% more likely to modify when describing complex (31%), compared to simple shapes (23%). The 95% confidence interval (*CI*) for the modification rate differences was $\pm 7.2\%$.[1] Note that the modification rate in the absence of a contrast was greater than that observed in other experiments, a point which we will return to later. Finally, trials that had a size contrast were included in a separate analysis of the distance between the target and the size contrast. The modification rate for trials with con-

[1] Throughout this paper, the 95% confidence intervals for paired comparisons are calculated using the *MSsxc* ANOVA term for that comparison. For interpretation of significant interactions, we calculate the 95% *CI* of the difference using the *MSsxc* term from a subsequent one-way ANOVA on difference scores (Masson & Loftus, 2003, p. 212). In the present case, the *MSsxc* was obtained from a subsequent ANOVA using presence of contrast as factor and participant difference scores (complex–simple) as dependent.

Table 1
Analysis of variance results for Experiment 1

Dependent	Factors	F_1[a]	p	F_2[a,b]	p	Min F'	p
Size adjectives (%)	Presence of size contrast (S)	$F(1,17) = 198.33$	<.0001	$F(1,16) = 1180.41$	<.0001	$F(1,23) = 169.80$	<.0001
	Complexity (C)	$F(1,17) = 1.50$	=.24	$F(1,16) = 1.21$	=.29	$F(1,32) = .67$	=.42
	S × C	$F(1,17) = 4.17$	=.06	$F(1,16) = 5.12$	<.05	$F(1,33) = 2.30$	=.14
Size adjectives (%): Contrast trials only	Display distance (D)	$F(1,17) = 4.50$	<.05	$F(1,16) = 1.62$	=.22	$F(1,26) = 1.19$	=.29
	Complexity (C)	$F(1,17) = 2.03$	=.17	$F(1,16) = .92$	=.35	$F(1,28) = .63$	=.43
	D × C	$F(1,17) = 3.10$	=.10	$F(1,16) = .90$	=.36	$F(1,25) = .70$	=.41
1st contrast fixation (ms)	Adjective position (A)	$F(1,9) = 172.17$	<.0001	$F(1,15) = 92.33$	<.0001	$F(1,24) = 60.1$	<.0001
	Complexity (C)	$F(1,9) = 51.19$	<.0001	$F(1,15) = 15.62$	<.01	$F(1,22) = 11.97$	<.01
	A × C	$F(1,9) = .97$	=.35	$F(1,15) = .12$	=.73	$F(1,18) = .11$	=.75
1st contrast fixation (ms): Pre-nominal NPs only	Display distance (D)	$F(1,14) = 4.85$	<.05	$F(1,16) = 1.07$	=.32	$F(1,23) = .88$	=.36
	Complexity (C)	$F(1,14) = 19.19$	<.001	$F(1,16) = 9.73$	<.01	$F(1,28) = 6.46$	<.05
	D × C	$F(1,14) = 2.89$	=.11	$F(1,16) = 2.02$	=.17	$F(1,30) = 1.19$	=.28
Speech onset latency	Adjective position (A)	$F(1,9) = 4.40$	=.07	$F(1,15) = 12.71$	<.01	$F(1,15) = 3.27$	=.09
	Complexity (C)	$F(1,9) = 1.84$	=.21	$F(1,15) = 1.39$	=.26	$F(1,24) = .79$	=.38
	A × C	$F(1,9) = 10.78$	<.01	$F(1,15) = 1.97$	=.18	$F(1,20) = 1.67$	=.21

[a] Fluctuations in the denominator *df* are due to missing cells.
[b] The complexity factor is treated as a between-items factor in the items analysis.

trasts that were closer in display distance was only 2% higher than the rate for trials in which the contrast was further away, an effect that was significant only in the participants analysis; the interaction was not significant.

For trials with a size contrast and a size adjective, the size term was used in a pre-nominal expression on 89% of the trials with simple shapes (e.g., *the small triangle*). For the complex shapes, the size term occurred in a post-nominal prepositional phrase on 73% of the trials (e.g., *the square with small triangles*). On a smaller proportion of trials, the size term occurred in a delayed, prosodically separate phrase that was typically associated with a disfluency, as in *the triangle*, *uh the small one*, or *the square with triangles*, *uh small ones*. For simple shapes, these post-noun repairs were used on 11% of trials when a contrast was present and a size modifier was used. For complex shapes, post-noun repairs occurred on 16% of these trials.

Fixation analyses

On approximately 13% of trials (74/587) with a size adjective, a size contrast in the scene, and a fixation to the size contrast, one or more of the speaker's first few words were disfluent, as in *thee small triangle*, or *thuuh square with small triangles*. Because there were relatively few of these trials, and because they were unevenly distributed across conditions and across participants, we restricted our analyses to trials where the onset of the utterance was fluent. An additional 11 trials were excluded from further analyses because the participant used the wrong size adjective. We examine disfluent utterances in Experiment 2, which uses a design that generates a higher proportion of disfluent trials.

To evaluate whether use of a size adjective depends upon whether the contrast was fixated by the speaker, we compared modification rates for trials with a contrast when the speaker did and did not look at the contrast. If speakers encode the size contrast when it is first fixated, then modification rates should decrease for trials where the speaker did not fixate on the contrast. This prediction was confirmed. When a size contrast was present in the scene, and speakers looked at the contrast, they used a size adjective on 99% of trials (502/508). When the size contrast was not fixated, the modification rate dropped to 68% (15/22). The 95% *CI* of the 30% difference in participant means was ±18.7%. The interaction of display distance and fixation on the contrast was not calculated due to a lack of trials without contrast fixations. In summary, the relationship between gaze and modification is consistent with our first hypothesis, that use of a size adjective depends in part on whether the speaker has fixated the size contrast, and allows us to evaluate our second, stronger hypothesis, which is that there will be a systematic relationship between the *timing* of the first look to the contrast and the form of the utterance.

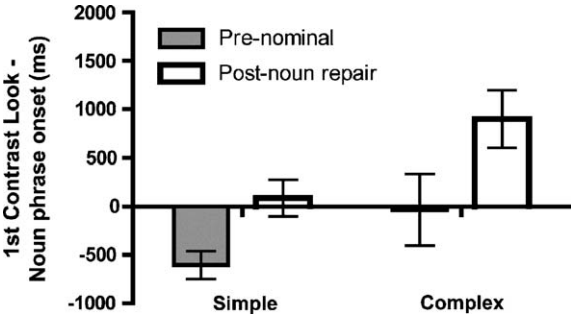

Fig. 2. Average first fixation times (first fixation to the size contrast, relative to noun phrase onset) for fluent referring expressions with pre-nominal size adjectives and post-noun size repairs, for both simple and complex shapes. Time is in milliseconds. Error bars indicate 95% *CI* of the mean, calculated independently for each group.

Fig. 2 shows the time of the first fixation on the contrast, relative to the onset of the utterance, hereafter "first fixation time," for trials with pre-nominal adjectives (e.g., *the small triangle*, or *the square with small triangles*) and post-noun repairs (e.g., *the triangle, uh small one*, or *the square with triangles*, *uh small ones*) for simple and complex shapes. The form of the referring expression is clearly related to the timing between the first look to the size contrast and the onset of the utterance.

Negative (early) first fixation times, which indicate that the first look preceded the onset of the utterance, are associated with use of pre-nominal adjectives. Positive (late) first fixation times, which indicate that the first look followed the onset of the utterance, are associated with utterances in which the adjective appeared in a post-noun repair.

In the following analyses, we examine the relationship between first fixation times, shape complexity, and adjective position. The display distance factor (distance between target and contrast) was not included to increase the number of participants and items included in each of our comparisons. A separate, planned analysis of display distance is presented at the end of this section.

For simple shapes with pre-nominal size adjectives, the mean first fixation was −605 ms compared to +88 ms for utterances with post-noun repairs. For complex shapes, the mean first fixation was −34 ms for utterances in which the adjective preceded the noun in the prepositional phrase, compared to +901 ms for post-noun repairs. An ANOVA revealed an effect of adjective position, with earlier first fixations for pre-nominal adjectives compared to post-noun repairs, and an effect of complexity, with earlier first fixations for references to simple shapes compared to complex shapes. The interaction between complexity and adjective position did not approach significance.

The shorter lags between the first fixation and the onset of the utterance for pre-nominal adjectives with complex shapes compared to pre-nominal adjectives with simple shapes suggests that new message elements *can* be added to a planned referential description during or immediately before production.[2] The delay in first contrast fixations for complex shapes was not disruptive because the size adjective was planned later in the phrase, allowing enough time for the adjective to be prepared and inserted into the utterance plan. Additionally, the delay in first fixations to the contrast when speakers placed size adjectives in a post-noun repair supports our second hypothesis that the timing of the first fixation to the contrast will be related the form of the utterance. When speakers fixated the contrast well before utterance onset, they were able to include a size adjective in a pre-nominal position, but when the contrast was not fixated until after utterance onset, size adjectives were included in a delayed post-noun repair phrase. This finding allows us to evaluate our third hypothesis about the size of the message that is passed on to utterance planning, and the timing between repairs to the message and utterance planning.

A closer inspection of the post-noun repairs indicates that the timing of the repairs was variable, with some repairs occurring earlier than others (e.g., *the rectangle...small rectangle*, vs. *the rectan-big rectangle*). The relationship between the delay in first fixating the contrast (relative to noun phrase onset), and the delay in the onset of the post-nominal size adjective (relative to the first contrast fixation) is presented separately for simple and complex shapes in Fig. 3. We chose to focus this analysis on post-noun repairs because the speaker is not under pressure to prepare additional words or phrases following the repair. Thus the timing of the repair was not likely to be influenced by subsequent planning processes. Additionally, only those trials with first fixations that occurred after noun phrase onset are included in the analysis.[3]

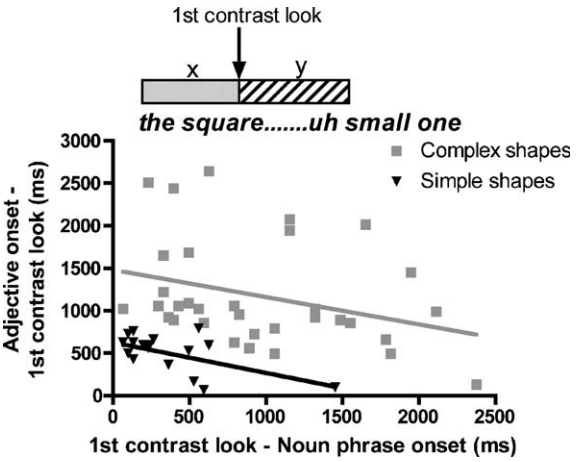

Fig. 3. Relationship between the first fixation time (first fixation to the size contrast, relative to noun phrase onset), and the timing of post-noun size adjectives, relative to first fixation time, for both simple and complex shapes. Only noun phrases with post-noun repairs, and positive first fixation times are shown. Time is in milliseconds.

Adjective delay (y axis) is the time between the first fixation and adjective onset; larger adjective delays mean a longer lag between the first fixation and the onset of the adjective. First fixation time is plotted on the x axis; larger values mean a longer lag between speech onset and the first fixation on the contrast. Values more than two standard deviations away from the mean were trimmed ($n = 4$). For simple shapes, we observed a negative relationship between first fixations and adjective onset time, $r^2 = .32$, $p < .05$ ($n = 19$); the effect was marginal for complex shapes, $r^2 = .10$, $p < .07$ ($n = 33$). These results tentatively suggest that when the need for a size adjective is first noticed after speech onset, preparation of the repair is delayed by preparation of the original utterance. Consistent with this interpretation is the additional finding that the lag between the first fixation to the contrast and the adjective onset was longer for complex shapes (mean = 1220 ms), compared to simple shapes (440 ms). One explanation is that speakers often included additional words following the noun inside the prepositional phrase when describing complex shapes (e.g., *the square with triangles in it... small ones*), but not with simple shapes (e.g., *the square...small one*). A remaining explanation, however, is that the size repair took longer to encode for the complex shapes.

Finally, the effect of display distance on first fixation times was analyzed in a separate, planned ANOVA which included only pre-nominally modified noun phrases; post-noun repairs were excluded due to a lack of data. The effect of display distance was significant in the participants analysis only, with earlier first fixations when the contrast was closer. The effect of shape

[2] Additional support for this conclusion comes from a small set of trials ($n = 23$) on which speakers ($n = 2$) used pre-nominal constructions to describe complex shapes, as in *the three small hearts in a square*. Here, the average first fixation time was −531 ms, mid-way between that for simple shapes with pre-nominal modifiers and complex shapes with size adjectives in a post-noun prepositional phrase. Unfortunately, due to the small number of participants who used this construction, we do not have enough statistical power to perform an analysis comparing these constructions with the post-nominal constructions.

[3] Inclusion of first fixations that occurred before noun phrase onset would violate assumptions of regression because the time between first fixation and noun phrase onset (x values) would be included in the time between first fixation and adjective onset (y values).

complexity was significant in this subset of the data, with faster first fixation times for simple shapes. The interaction did not approach significance, suggesting that small delays in fixating the contrast when it was further away may not have been long enough to disrupt planning of the adjective.

Time to begin speaking

Added evidence that messages can be passed along in units smaller than a full referential description comes from a comparison of the delay between the onset of the display and the time to begin speaking for simple and complex shapes. Fig. 4 shows the average onset of fluent referring expressions for simple and complex shapes, relative to display onset. For simple shapes, when participants used a pre-nominal modifier, as in *the small triangle*, the referring expression began an average of 1546 ms after the display appeared compared to 1061 ms for post-noun repairs such as *the triangle, uh small one*.

However, for complex shapes, speech onsets were only slightly faster when a post-noun repair was used. When describing a complex shape with a pre-nominal modifier, speakers began their referring expressions an average of 1434 ms following display onset compared to 1390 ms for utterances that contained a post-noun repair (e.g., *the square with triangles, uh small ones*). An ANOVA using utterance onset lag as the dependent measure revealed a significant effect of adjective position (marginal by participants), no effect of complexity, and an interaction between complexity and adjective position that was significant in the participants analysis only. The 485 ms delay in speech onset times for simple shapes with pre-nominal adjectives was reliable, but the 44 ms delay for complex shapes with pre-nominal modifiers was not, 95% *CI* of the difference = ±205 ms.

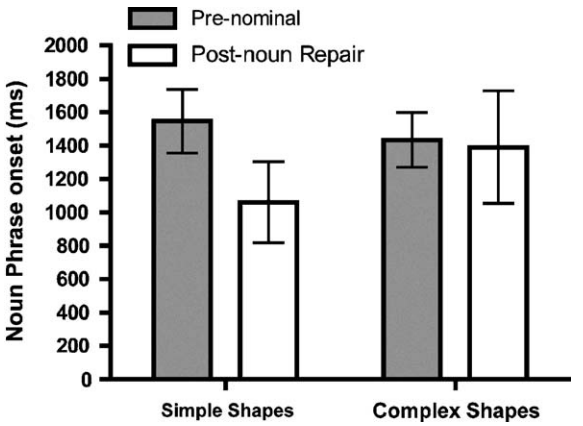

Fig. 4. Average noun phrase onset relative to display onset, for fluent referring expressions with pre-nominal size adjectives or post-noun size repairs for both simple and complex shapes. Time is in milliseconds. Error bars indicate 95% *CI* of the mean, calculated independently for each group.

For simple shapes, the longer speech onset times for referring expressions with pre-nominal adjectives compared to those with post-noun repairs, suggests that for the post-noun repairs, speakers did not begin preparation of the size adjective before starting to speak. This finding is consistent with the results from the analysis of eye movements and supports the hypothesis that sub-reference sized message units can be prepared and passed to utterance planning individually. In contrast, for complex shapes, adjectives that preceded the noun in the prepositional phrase were not associated with delayed speech onsets compared to post-noun repairs. This finding adds to the evidence from the eye movement record that for complex shapes, speakers did not prepare pre-nominal size adjectives or post-noun size repairs until after utterance onset.

Discussion

The results demonstrate a clear relationship between initial eye movements to message-relevant, but non-mentioned entities, and the form of the utterance. Modification rates were modulated by the presence of a size contrast, replicating previous findings by Sedivy (2001, 2003) using simpler displays. In addition, modification rates were strongly affected by whether or not the speaker fixated the size contrast. Thus the first look to the contrast can be used to infer when the need for including size in the message was first encoded. Most importantly, the timing of the first contrast fixation was strongly linked to the form of the utterance. When speakers used a pre-nominal adjective, they typically fixated the size contrast more than 600 ms before the onset of the utterance. When speakers placed the size term in a prosodically separate phrase that occurred after the noun that was being modified, they typically fixated the contrast after the onset of the utterance. Thus, fixations on the size contrast predict use of a size adjective, and the timing of these fixations predicts the form of the utterance.

Differences in timing of first contrast fixations for descriptions of simple and complex shapes provide insights into the size of message planning units, and the mechanisms for repairing messages and utterances. When speakers used a pre-nominal adjective, fixations to the contrast occurred later with respect to the onset of the utterance for complex phrases compared to simple phrases. The delay in fixations to the contrast for complex shapes compared to simple shapes suggests that speakers were able to prepare the size term later for the complex shapes. This result suggests that when preparing a modified referring expression, messages are incrementally prepared and passed onto utterance planning in units smaller than the size needed for an entire referring expression. Additionally, for first fixations that occurred after speech onset, the earlier the fixation, the longer the delays in use of a post-noun repair. This result

suggests that when messages are updated to include size, repair to the utterance plan may sometimes be delayed by processing of the original utterance plan.

Two aspects of the results are unexpected. First, the 600 ms lag between the fixation to the contrast and the onset of the utterance was several hundred milliseconds faster than one might expect, given other results in the literature that find about an 800 ms lag between the fixation to a referent and the onset of its name. Second, size adjectives were used on more than 25% of trials when there was not a contrast. This is much higher than the ~5% modification rate reported in previous studies in the absence of contrast (Sedivy, 2001). One possible explanation for these results is that a small number of shapes were repeatedly used on target trials, which might reduce the demands of lexical encoding, and 25% of the trials had targets with size contrasts, which might make inclusion of size in the description somewhat formulaic. Thus, message and utterance planning times might not be representative of those that would occur with more demanding lexical encoding and a lower proportion of trials with contrast. Experiment 2 was designed to replicate the main results from Experiment 1, while increasing the number of target types and decreasing the proportion of target trials with size contrasts.

Experiment 2

This experiment used displays containing pictures of common objects. This allowed us to increase the set of target names from the nine basic shapes used in Experiment 1 to more than nine hundred objects with different names. We also increased the length of the experiment to three hours, which allowed us to reduce the proportion of target trials with size contrasts from 25 to 18%.

An important consequence of having more trials, and a larger set of target names, is that speakers are likely to produce more disfluent noun phrases. Fixations to the contrast associated with these utterances have the potential to provide a window into the interplay between message updating and utterance planning. For example, Ferreira and Dell (2000) have argued that speakers insert optional words such as *that* to buy time to complete lexical retrieval for an upcoming word (also see Jaeger & Wasow, 2005). Similarly, Arnold and colleagues (Arnold, Fagnano, & Tanenhaus, 2003; Arnold & Tanenhaus, in press) have demonstrated that disfluencies are more likely to precede a word that is new to the discourse compared to a word that is old, and therefore easier to produce. Speakers might also use disfluencies to buy time to include an updated element of the message into the current phrase. For example, if a contrast is first encountered shortly after utterance planning has begun, then elongating the determiner, pausing, or uttering a filler morpheme might allow size to be incorporated into the referring expression as a pre-nominal modifier rather than as a repair in a second phrase. If this is the case, then, whereas early looks to the contrast are likely to result in a fluent pre-nominal utterance and late looks a post-noun repair, intermediate looks would result in a disfluent pre-nominal utterance.

This experiment allows us to investigate three hypotheses, the first two of which were outlined in the introduction: first, fixations to the contrast should predict whether or not the speaker uses a size modifier, with higher rates of modification when the contrast has been fixated. Second, the timing of the first fixation to the contrast with respect to the onset of the utterance should predict whether the speaker uses a pre-nominal adjective or an adjective in a post-noun repair phrase, with shorter lags associated with pre-nominal modification. Third, intermediate delays in fixating the contrast should result in disfluent, pre-nominally modified utterances.

Method

Participants

Twenty pairs of participants who were members of the undergraduate community at the University of Rochester were paid for their participation. All participants were native speakers of North American English, and all partners self-identified as being friends.

Procedure

The equipment and set-up were the same as in Experiment 1; both participants had microphones and one wore a head-mounted ASL brand eye-tracker. The voices of both partners, as well as a record of the eye movements, superimposed on a video-record of the scene, were recorded to digital video at 30 Hz. Because of the length of the experiment, participants completed the experiment in three 1-h sessions. The sessions were distributed over the course of two or three days, depending on availability.

Participants took turns speaking for a total of 240 trials. Each trial featured a different set of 14 objects, some of which differed in size, e.g., a small horse and a large horse. Fig. 5 illustrates an example screen from an experimental trial. Each picture was a simple object, selected to be easily identifiable and nameable by participants. A total of 983 different pictures were used across the 240 trials; each picture was presented an average of 3.42 times. No picture was presented more than seven times.

On each trial, 14 objects were arranged on the screen in two distinct areas, separated by a 'river'. The scenes also featured a house, which was situated in the river. There were two target referents on each trial. The initial screens for the two participants were identical, but the objects were arranged in such a way to minimize the

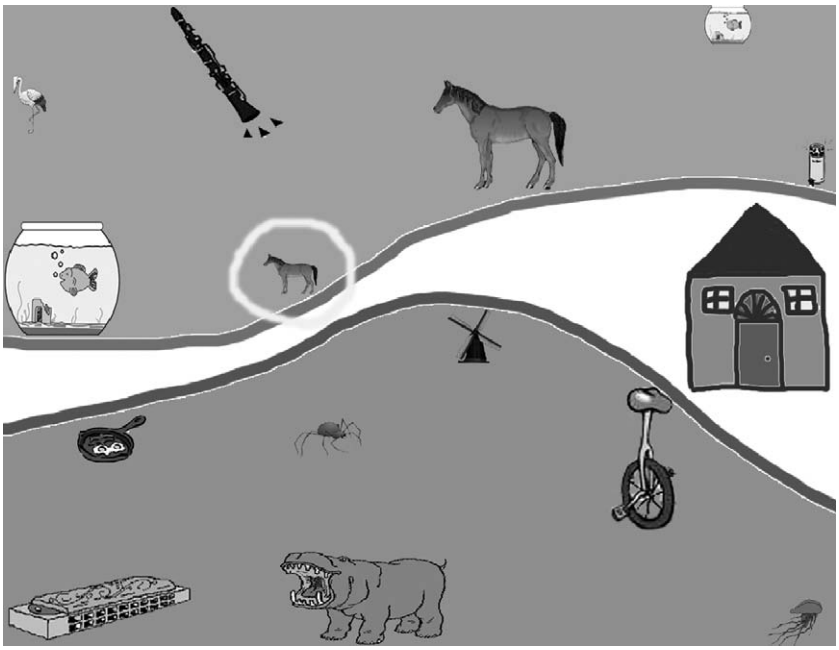

Fig. 5. Example stimulus display, shown from the speaker's perspective. On the listener's screen, the target is not circled but the display is otherwise identical.

use of locative terms. On each trial, both participants pressed a key to see the initial scene. The speaker then pressed a key a second time. On half of the trials, after this second mouse click a yellow line appeared around one of the two areas of the screen, as an indication that the first target referent would be in this area. On the other half of trials the screen remained unchanged. The speaker then pressed a key a third time, and for all trials, the third key-press caused a yellow circle to appear around the target referent. The speaker then instructed her partner to click on the target referent. After he clicked on the picture it moved to the house, and then off the screen with a second mouse-click. The speaker then received a second prompt marking a new target picture. After her partner clicked on this picture, a new trial began and the partner became the speaker. Each new trial began with a different array of pictures.

The structure of the game allowed for a high degree of interactivity, while preserving a trial structure that allowed us to control the referents and the referential context. The large number of trials and target referents allowed us to examine the production of size adjectives without having a large number of trials that required modification. Of the 480 referents across the 240 trials, only 86 were displayed with a size contrast (18%). While we collected eye movement data from the eye-tracked partner when they were speaking and when they were listening, we focus exclusively on a sub-set of trials during which the eye-tracked participant was speaking.

Results and discussion

Selection of trials

Each eye-tracked participant produced a total of 240 referring expressions. To reduce the amount of data to be analyzed, and increase the homogeneity of the referring expressions to be analyzed, we selected a subset of these trials for further analysis. Because producing the first reference may have affected the production of the second, we examined only the first referring expression on each trial. We then analyzed only trials on which the target referent was not in a contrast set with any other item on the screen, and trials on which the target referent had a size contrast in the same area (e.g., a small and a large horse on the same side of the river). We will refer to these as 'no-contrast' trials, and 'size contrast' trials, respectively. The same area constraint excluded trials in which the size contrast was on the opposite side of the river as the target. We eliminated those trials because speakers often started their utterance by specifying that the target was above or below the river, thus eliminating the need for modification when the target and contrast were on different sides.

To select a homogenous set of trials, we conducted a post hoc norming study to establish the name-agreement and disfluency rates for target pictures from the remaining 104 trials. Twelve participants who did not participate in Experiments 1 or 2 participated in this rating study. Each participant was tested separately, and self-

identified as a native speaker of North American English. The study took approximately fifteen minutes; participants were paid for their time.

Target pictures were presented one at a time on a computer screen. Participants were asked to name each picture. Each picture was described once, resulting in a total of 104 trials. A digital Hi-8 camcorder recorded the participants' responses, which were later transcribed.

Name agreement was quantified as the proportion of trials on which the 12 participants used the same name to describe the picture (Griffin & Huitema, 1999; Snodgrass & Yuditsky, 1996). References that used different determiners such as *the horse* and *a horse* were considered to be tokens of the same name; phrases that were modified differently, or used a different head noun, such as *the brown horse* and *the donkey* were considered to be tokens of a different name. A referring expression was considered to be disfluent if it contained a pause, a repair, a repeat, a lengthened word, or a filled pause such as *um* or *uh*. The data were used to select 30 contrast and 30 no-contrast pictures which were matched for name agreement (74%) and disfluency rates (18%).

In the main experiment, each of the 20 participants generated a trial for each of the 60 pictures, resulting in a total of 1200 trials. Of these trials, 63 (5%) were excluded from further analysis either because the speaker skipped the trial, did not refer to the object (e.g., *uh*, *I don't know*), or because of equipment failure, leaving a total of 1137 trials (567 no-contrast, and 570 contrast trials).

Referential form

Speakers tailored their messages to the referential context, rarely using a size adjective when there was not a contrast in the display. Speakers used size adjectives for only 1% of references when there was not a size contrast in the display (as compared to 27% for simple shapes in Experiment 1). Speakers used a size adjective for 72% of the references on trials with a contrast, a significantly higher modification rate than for those trials without a contrast present, 95% *CI* of the difference $= \pm 6.6\%$. These proportions are consistent with those found by Sedivy and colleagues (Gregory et al., 2003; Sedivy, 2003, 2001), in experiments that used simpler scenes with fewer objects. On the 413 size contrast trials where speakers used a size adjective, 62% were pre-nominally modified, e.g., *the small horse*, and 37% had post-noun repairs, e.g., *the horse*, *OH the small one*.

Fixation analyses

Analysis of eye fixations was restricted to trials were there was a size contrast on the same side of the river as the target. Fluent ($n = 369$) and initially disfluent trials ($n = 201$) are analyzed separately. Because we were primarily interested in the planning of the size adjective, trials were considered disfluent only if they contained a disfluency *before* the head noun or pre-nominal size adjective (e.g., *thee small horse*, *the. . .small horse*); trials with a disfluency after the head noun or pre-nominal size adjective were grouped with the fluent trials (e.g., *the small uh horse*; *the horse. . .small one*). The speaker's eye movements were analyzed from the time the target referent was highlighted, until 2000 ms after speech offset, or the beginning of the next trial, whichever came first.

Fluent trials. As in Experiment 1, and consistent with our first hypothesis, looks to the contrast strongly predicted whether the size of the target would be mentioned. On 86% of the fluent trials during which the speakers looked at the contrast, the referring expression included a size modifier (226/262). When speakers did not look at the size contrast, they used a size modifier on only 19% of trials (20/107). The difference in modification rate based on participant means was 62%, a significant difference, 95% *CI* of the difference $= \pm 8.5\%$. Note, however, that the 19% modification rate for trials without a fixation to the contrast picture was greater than the 1% modification rate for trials without a contrast in the display. This suggests that, on some trials, the contrast had been coded prior to the button press that highlighted the target.

As in Experiment 1, and consistent with our second hypothesis, the first fixation to the contrast for the fluent trials was systematically related to the form of the utterance (see Fig. 6). For fluent utterances with pre-nominal adjectives, e.g., *the small horse*, the mean first fixation to the contrast was −887 ms, before the onset of the refer-

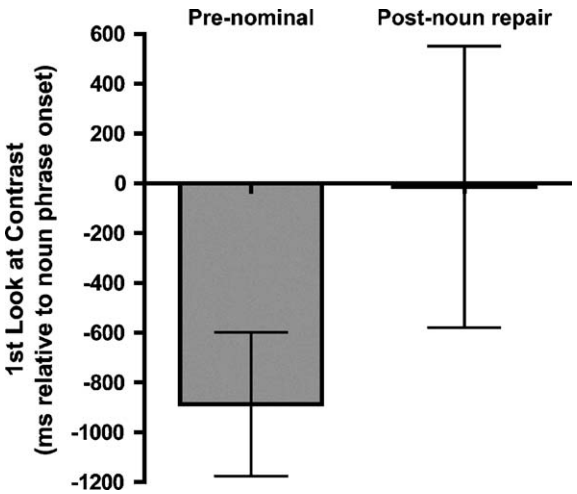

Fig. 6. Average first fixation times (first fixation to the size contrast, relative to noun phrase onset) for fluent referring expressions with pre-nominal size adjectives and post-noun size repairs. Time is in milliseconds. Error bars indicate 95% *CI* of the mean, calculated independently for each group.

ring expression, whereas for utterances with repairs, the mean first fixation was significantly later, −14 ms. This 873 ms difference in first fixation times was significant, 95% *CI* of the difference = ±437 ms.

The 873 ms difference in first fixation time between references with pre-nominal modification and references with post-nominal modification is similar to the 693 ms difference found in Experiment 1. However, in this experiment, fixations to the contrast occurred earlier relative to the onset of the utterance by approximately 100–200 ms. For pre-nominal adjectives in Experiment 2, the first fixation was −887 ms, compared to −605 ms in Experiment 1. For post-noun repairs, the mean first fixation in Experiment 2 was −14 ms compared to +88 ms for Experiment 1. Recall our concern from Experiment 1 that the repeated use of shapes reduced planning time. The results from this experiment are more in line with what we would expect on the basis of previous studies in which fixations to a referent precede the onset of speech by 800 ms or more.

Disfluent trials. In the 201 disfluent trials that we analyzed, speakers elongated the initial determiner and included a pause or a filled pause, such as *theee uh small horse*. As with the fluent trials, when speakers fixated the contrast, they were significantly more likely to use a size adjective (91%) than if they had not (34%), 95% *CI* of the difference = ±16.3%.

For those disfluent trials with at least one fixation to the size contrast, the first fixation to the contrast showed the same relationship to utterance form as the fluent utterances. First fixations were on average −528 ms for utterances with pre-nominal modifiers compared to +819 ms for post-noun repairs. This 1347 ms difference in first fixation time was significant, 95% *CI* = ±713 ms. Note, however, that for pre-nominally modified utterances, the first fixation was over 350 ms later for the disfluent utterances compared to the fluent utterances, a reliable difference, 95% *CI* of the difference = ±161 ms. This difference in first fixation times for disfluent compared to fluent noun phrases suggests that more than 500–900 ms of planning time is typically required to revise a message without disturbing ongoing speech.

The relationship between the timing of the first look to the contrast and utterance form is highlighted in Fig. 7, which shows a frequency distribution of first contrast fixations for fluent pre-nominally modified references, disfluent pre-nominally modified references, and fluent post-noun repairs.

The earliest fixations are associated with a fluent pre-nominal utterance whereas the latest fixations are associated with a post-noun repair. Intermediate fixations are associated with a disfluent utterance with a pre-nominal adjective. This finding is consistent with our third hypothesis, and suggests that the speaker was able to use disfluency to buy time to include size in the noun phrase as a pre-nominal adjective rather than as a post-noun repair.

As in Experiment 1, we observed a relationship between the time speakers first fixated the size contrast, and the *timing* of the size adjective for post-noun repairs (see Fig. 8). This analysis includes only those trials in which the speaker used a post-noun repair, and the first fixation on the contrast occurred after noun phrase onset. Additionally, data points more than two standard deviations away from the mean were trimmed ($n = 9$). Replicating the results for references to simple shapes in Experiment 1, when speakers were fluent, the delay in first fixating the size-contrast (relative to noun phrase onset) had a significant negative relationship with the delay of the post-noun repair adjective (relative to the first fixation), $r^2 = .13$, $p < .05$ ($n = 45$). The delay in first

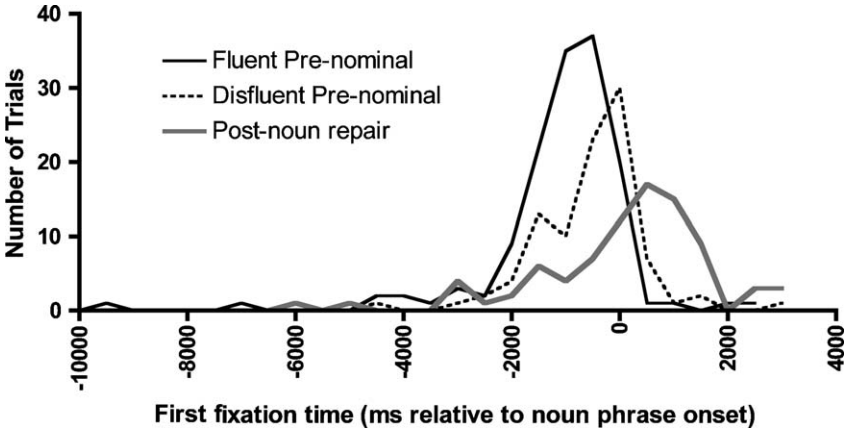

Fig. 7. Frequency distribution of the onset (in milliseconds) of the first eye movement to the contrast item (e.g., large horse), relative to the onset of the noun phrase (e.g., *the* in *the small horse*). Solid line indicates fluent, pre-nominally modified references like *the small horse*; dotted line indicates disfluent, pre-nominally modified utterances like *thee uh small horse*; grey line indicates initially fluent utterances with post-noun repairs such as *the horse. . .uh the BIG one*. 0 ms = noun phrase onset; bin size = 500 ms.

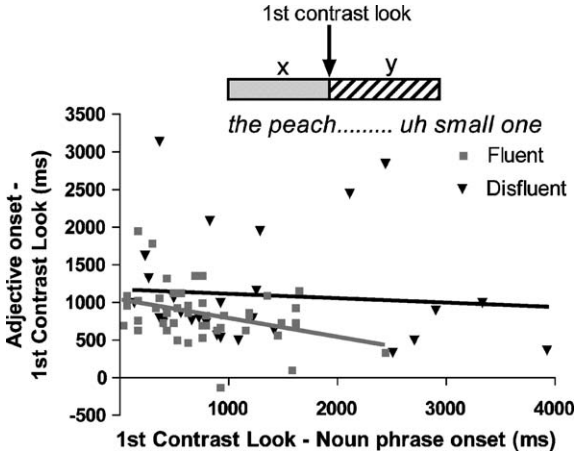

Fig. 8. Relationship between the first fixation time (first fixation to the size contrast, relative to noun phrase onset), and the timing of post-noun size adjectives, relative to first fixation time for both fluent and disfluent referring expressions. Only noun phrases with post-noun repairs, and positive first fixation times are shown. Time is in milliseconds.

fixations did not significantly affect adjective delay for disfluent utterances, $r^2 = .01$, $p = .67$ ($n = 29$).

The observed relationship for fluent forms adds to the evidence that formulation and preparation of repairs may be delayed by preparation of the original utterance. Increased variability associated with speaker difficulty may explain why this effect did not reach significance for disfluent forms.

Time to begin speaking

On each trial, the speaker pushed a button to highlight the target object with a yellow circle. The average time between the highlighting of the target and the noun phrase onset was slightly less than 2 s. However, unlike Experiment 1, the utterance onset time was not significantly affected by the form of the utterance. For initially fluent utterances, speakers started pre-nominally modified expressions 345 ms faster than expressions with post-noun repairs, a difference that was not reliable, 95% *CI* of the difference = ±259 ms. For disfluent utterances, pre-nominally modified expressions were only 120 ms faster than utterances with post-noun repairs, 95% *CI* of the difference = ±258 ms. Finally, speakers uttered disfluent, pre-nominally modified expressions, on average 66 ms faster than fluent, pre-nominally modified utterances, but this difference was also not reliable, 95% *CI* of the difference = ±201 ms.

General discussion

Three primary results emerged from our experiments. First, speakers typically used size adjectives in a referential expression only when there was a similar object in the referential domain that differed in size. This result provides additional evidence in support of Sedivy and colleagues' claims about the conditions under which speakers use scalar adjectives (Gregory et al., 2003; Sedivy, 2001, 2003). Second, use of a scalar adjective was significantly more likely if the speaker had fixated on the size contrast. This result demonstrates that the first fixation to the contrast provides a reasonable estimate of when the speaker first noticed the contrast object in our displays. These two results were prerequisites for exploring relationships between the timing of the first fixation to the contrast and the form of the referring expression that could be used to make inferences about the interface between message formulation and utterance planning.

Our third, and most important result is that the timing of the first fixation to the contrast was indeed related to the form of the referential expression. Earlier fixations to the contrast were associated with the use of a pre-nominal adjective. Later first fixations were associated with a post-noun repair that modified the description to add information about size. In addition, first fixations occurred later, with respect to the onset of the utterance, for referential descriptions in which the adjective occurred in a prepositional phrase, which modified a noun phrase, compared to referential descriptions with a pre-nominal adjective. Finally, disfluent utterances with pre-nominal adjectives were associated with first fixations to the contrast that were intermediate between first fixations associated with fluent use of a pre-nominal adjective, and first fixations associated with a post-noun repair.

These results provide preliminary support for several hypotheses about the interplay between message formulation and utterance planning. First, when an error is detected in the utterance plan after speaking has begun, initiation of a repair may be delayed by preparation of the original utterance. Repair phrases occurred at various points in time following the initial phrase, and for those trials when the contrast was noticed late, the timing of repair phrases was a negative function of the time that the contrast was first fixated. This result suggests that initiation of a repair to an utterance plan may sometimes be delayed until after the original utterance plan is complete.

Second, messages that will be mapped onto referring expressions can be constructed and passed onto utterance planning incrementally. The delay in first contrast fixations for pre-nominal expressions when describing complex shapes compared to simple shapes in Experiment 1 suggests that speakers were able to delay planning of the size adjective when it was to occur later in the referring expression. We can imagine two possible mechanisms for this delay. When the contrast is noticed earlier, speakers may access the typical syntactic frame

for the construction and delay accessing the scalar word form because it occurs relatively late in the syntactic frame. However, this possibility would not explain the delay in average first-contrast fixations for complex shapes. A second possibility is that when the speaker intends to refer to a complex shape and the contrast is noticed late, the utterance plan can be revised to include a size adjective. In this case, planning of the initial part of the utterance would not be disrupted (e.g., with a disfluency) because description of a complex shape allows the speaker to place the size modifier late in the referring expression, providing extra time for the repair. In contrast, simple shapes require early placement of size modifiers, and as a result, if the size contrast is noticed late, the speaker must use a post-noun repair. Thus, differences between complex and simple shapes in the canonical position of size adjectives would delay the mean first fixation time for pre-nominally modified complex shapes compared to simple shapes. These results suggest that for complex shapes, the size information can be planned separately from the first phrase.

Added support for this hypothesis comes from the analysis of disfluent, pre-nominally modified expressions in Experiment 2. Speakers who encountered information that size modification was necessary only a few hundred milliseconds before noun phrase onset sometimes elongated initial components of the noun phrase, which created a delay in the articulation process which allowed just enough time to insert a size adjective into the original phrase. This suggests that pre-nominal size information can be prepared separately from other components of a simple noun phrase. This finding also suggests that repair processes can be initiated without delay if an error in the message is detected before utterance onset.

In addition, the results suggest that just as speakers can use disfluency to buy time to plan an upcoming word or phrase, speakers can also use disfluency to provide enough time to add additional information to a planned utterance, based on a revised message. Added support comes from work by Ferreira and Swets (2002) who compared utterance durations in two experiments that differed in the amount of preparation time given to speakers. When speakers had less time to prepare their sentences, they spoke more slowly. These findings suggest that speakers may be able to control aspects of production, including disfluency and speech rate to accommodate various speaking pressures (see Schriefers & Teruel, 1999; Deese, 1984). An alternative interpretation of our results, however, is that the disfluency was due to something unrelated to the need to incorporate the size adjective. If this was the case, the disfluency was accidental rather than strategic. Nonetheless, it still would have provided the speaker with enough additional planning time to allow the size adjective to be planned and uttered pre-nominally rather than in a second phrase. We hope to distinguish between these alternatives in future work by manipulating the time at which speakers first notice the size contrast.

In work related to our analysis of disfluency, Griffin (2004a) analyzes a corpus of word substitution errors (e.g., *the hor-uh donkey*) made during object naming tasks. While our finding that disfluencies follow intermediate delays in gazing at the contrast might suggest that substitution errors should be associated with delayed gaze at targets, the opposite was found: Speakers gazed longer before name onset for erroneous names compared to correct names. In the present work, we focus on disfluent productions of the initial portions of noun phrases, including elongations and filler words (e.g., *thee uh horse*), whereas Griffin focuses on substitution errors on the object name. An important question for future work will be to clarify how different types of non-fluent speech index different aspects of planning difficulty (e.g., grammatical encoding vs. lexical retrieval). Another distinction between the two experiments is that our analysis focuses on the timing of the initial fixation on a non-mentioned item, whereas in Griffin (2004a), the measured gaze is the final fixation on the named object before articulation. Perhaps, for reasons to be explored in future work, difficulty in production is more likely to be reflected in the timing of earlier gazes, or gaze at message-relevant, but non-mentioned objects.

In summary, new visual information encountered during language production led speakers to repair their utterances—a process that requires continuous interaction between message formulation processes and the processes that govern utterance generation. These results place constraints on models of the interplay between message formulation and utterance planning. They also demonstrate that it is possible to examine message formulation by combining eye movement measures with studies of non-scripted, task-oriented dialogue.

Appendix

Experiment 1 Stimulus materials

Simple shapes: Circle, triangle, star, square, rectangle, oval, moon, heart, diamond.

Complex shapes: Squares, rectangles, and circles which contain three triangles, hearts or stars, for a total of nine complex shapes.

References

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research, 32*, 25–36.

Arnold, J. E., & Tanenhaus, M. K. (in press). Disfluency isn't just um and uh: The role of prosody in the comprehension of disfluency. In E. Gibson & N. Pearlmutter (Eds.), The

processing and acquisition of reference. Cambridge, MA: MIT press.

Blumenthal, A. L. (1970). *Language and psychology: Historical aspects of psycholinguistics*. New York, NY: John Wiley & Sons, Inc..

Bock, J. K. (1995). Sentence production: From mind to mouth. In J. Miller & P. Eimas (Eds.), *Handbook of perception and cognition: Vol. 11. speech, language, and communication* (pp. 181–216). New York, NY: Academic Press.

Bock, J. K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language, 48*, 653–685.

Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2005). Real-time reference resolution by naïve participants during a task-based unscripted conversation. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language as product and language as action traditions* (pp. 153–171). Cambridge, MA: MIT press.

Clark, H. H. (1991). Words, the world, and their possibilities. In G. Lockhead & J. Pomerantz (Eds.), *The perception of structure* (pp. 263–277). Washington, D.C: APA.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84–107.

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 & Language, 46*, 178–198.

Deese, J. (1984). *Thought into speech: The psychology of a language*. Englewood Cliffs, NJ: Prentice-Hall.

Dell, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review, 93*, 283–321.

Dell, G. S., & O'Seaghdha (1992). Stages of lexical access in language production. *Cognition, 42*, 287–314.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory & Language, 30*, 210–233.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*, 296–340.

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.

Gregory, M. L., Joshi, A., Grodner, D., & Sedivy, J. C. (2003, March). *Adjectives and processing effort: So, uh, what are we doing during disfluencies?* Paper presented at the 16th Annual CUNY Conference on Human Sentence Processing, Cambridge, MA.

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

Griffin, Z. M. (2004a). The eyes are right when the mouth is wrong. *Psychological Science, 15*, 814–821.

Griffin, Z. M. (2004b). Why look. In F. Ferreira & J. M. Henderson (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 213–247). New York, NY: Psychology Press.

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

Griffin, Z. M., & Huitema, J. S. (1999). Beckman spoken picture naming norms, Retrieved January 1st, 2004 from the University of Illinois, Beckman Institute <http://langprod.cogsci.uiuc.edu/~norms/>.

Henderson, J. M., & Ferreira, F. (Eds.). (2004). *The interface of language, vision, and action: Eye movements and the visual world*. New York, NY: Psychology Press.

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

Jaeger, T. F., & Wasow, T. (2005, March). Production complexity driven variation: The case of relativizer distribution in non-subject-extracted relative clauses. Paper presented at the 18th Annual CUNY Conference on Human Sentence Processing, Tuscon, AZ.

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

Levelt, W. J. M., & Maassen, B. (1981). Lexical search and order of mention in sentence production. In W. Klein & W. J. M. Levelt (Eds.), *Crossing the boundaries in linguistics* (pp. 221–252). Dordrecht: Reidel.

Levelt, W. J. M., Roelofs, A. P. A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–37.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 203–220.

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analysis: Achievements, limitations, and alternatives. *Cognition, 42*, 181–212.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory & Language, 35*, 477–496.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition, 89*, 25–41.

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review, 77*, 257–273.

Osgood, C. E. (1971). Where do sentences come from? In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*. Cambridge, MA: Cambridge University Press.

Paul, H. (1880). *Prinzipien der Sprachgeschichte*. Leipzig: Niemeyer, [Principles of the history of language].

Pickering, M. J., & Garrod, S. C. (2004). Towards a mechanistic theory of dialog. *Behavioral and Brain Sciences, 7*, 169–190.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition, 76*, 269–295.

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.

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

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research, 32*, 3–23.

Sedivy, J. C. (2001, March). *Evidence of Gricean expectations in on-line referential processing.* Paper presented at the 14th Annual CUNY Conference on Human Sentence Processing, Philadelphia, PA.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavioral Research Methods, & Instruments Computers, 28*, 516–536.

Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica, 45*, 175–197.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Trueswell, J. C., & Tanenhaus, M. K. (Eds.). (2005). *Approaches to studying world-situated language use: Bridging the language as product and language as action traditions*. Cambridge, MA: MIT Press.

van der Meulen, F. F. (2001). Moving eyes and naming objects. Unpublished doctoral dissertation, Katholieke Universiteit Nijmegen, Nijmegen, The Netherlands.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 92–114.

Wundt, W. (1900). Völkerpsychologie: Vol. 1–2. Die Sprache [Language]. Leipzig: Kröner-Engelmann.