Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 54 (2006) 554–573
Journal of Memory and Language
www.elsevier.com/locate/jml

Do speakers and listeners observe the Gricean Maxim of Quantity?

Paul E. Engelhardt [a,*], Karl G.D. Bailey [b], Fernanda Ferreira [a]

[a] *Department of Psychology, Michigan State University, East Lansing, MI 48824-1116, USA*
[b] *Andrews University, Berrien Springs, MI 49104, USA*

Received 25 April 2005; revision received 12 December 2005
Available online 21 February 2006

Abstract

The Gricean Maxim of Quantity is believed to govern linguistic performance. Speakers are assumed to provide as much information as required for referent identification and no more, and listeners are believed to expect unambiguous but concise descriptions. In three experiments we examined the extent to which naïve participants are sensitive to the Maxim of Quantity. The first was a production experiment which demonstrated that speakers over-describe almost one-third of the time. The second experiment showed that listeners do not judge over-descriptions to be any worse than concise expressions. The third experiment used the Visual World Paradigm to assess listeners' moment-by-moment interpretations of over-described utterances. This last experiment revealed that over-descriptions trigger eye movements that can be interpreted as indicating confusion. The results provide support for the use of a simple heuristic such as Minimal Attachment or Argument Saturation to create an initial parse. We conclude that people are only moderately Gricean.

Keywords: Gricean Maxim of Quantity; Syntactic ambiguity resolution; Visual World Paradigm; Language comprehension; Eye movements; Language production

A simple object such as an apple may be described as *the apple*, *the ripe apple*, *the ripe apple on the towel*, *the ripe apple on the towel in the kitchen*, and so on. How does a speaker decide how much information to provide when referring to an object? Psycholinguists generally assume that speakers are cooperative and adhere to a Maxim of Quantity, which consists of two principles (Grice, 1975). The first is that speakers should make their contributions as informative as required: enough information should be included to allow an addressee to identify an intended referent. If there is only one apple in the relevant environment, then *the apple* is sufficient. If there is more than one, enough modification should be provided to allow the apple to be uniquely identified. We will use the term *under-description* to refer to any expressions not consistent with this principle. The second component of the Maxim of Quantity is that speakers should not make contributions more informative than is necessary for successful communication. If there is only one apple, modifiers are not required and should not be included. We will use the term *over-description* to refer to expressions that are inconsistent with this second component of the Maxim of Quantity.

* Corresponding author. Fax: +1 517 353 3745.
E-mail address: paul@eyelab.msu.edu (P.E. Engelhardt).

0749-596X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2005.12.009

It is easy to see why this Gricean principle has had such an enduring appeal. Certainly, if a speaker wishes to be understood, it makes sense that he or she would avoid producing expressions that do not contain enough information to allow a listener to identify the intended referent. Similarly, it seems intuitively obvious that speakers would want to minimize their efforts and not include a potentially infinite number of modifiers that are not likely to be informationally useful. Because it seems plausible that speakers would adhere to the Maxim of Quantity, and because it seems reasonable to assume that conversational partners are cooperative (Clark, 1992), listeners will have certain expectations of speakers. Listeners expect them to provide enough information to allow referent identification, and will therefore be bothered by under-descriptions. In addition, they expect all information provided to be relevant, so if a modifier is included, it should be informative (in the technical sense of ruling out states of affairs; Levison, 2000). Thus, over-descriptions are more than just a waste of time; they may actually disrupt comprehension. For example, if a speaker uses the expression *the large apple*, the listener will assume that *large* is communicatively relevant, and so she might assume that there is more than one apple in the discourse and that the modifier is the information allowing the correct apple to be distinguished from any others. If in fact only one apple exists, the listener has been misled about the existence of a set of objects, which may impair communication (cf. Sedivy, Tanenhaus, Chambers, & Carlson, 1999).

While these ideas are intuitively appealing, it is important to know whether people adhere to these principles of communication in practice. What is the evidence that the Maxim of Quantity is followed, both by speakers and listeners? The experiments described below were designed to answer this question. Previous work by Deutsch and Pechmann (1982) examined this issue by using contexts that contained multiple objects of the same type. Objects differed in size and color, and participants were asked to name one object that they would like to give as a present to someone else. In contexts with multiple referents, Deutsch and Pechmann found that adult speakers rarely produced ambiguous or under-described utterances. For example, in a display containing a red apple and a green apple, speakers almost never used the expression *the apple*. Instead, they produced enough modification to allow the intended referent to be identified. Surprisingly, redundant or over-described utterances were produced fairly often, on over one-quarter of trials. For example, in a display containing a single apple, participants would often refer to it as *the red apple* or *the large apple*, although the phrase *the apple* was sufficient for identification. Moreover, they argued that these over-described utterances actually led to more efficient searches, indicating that comprehenders do not suffer when they encounter over-descriptions. The findings that speakers over-describe and that listeners perform efficiently with over-descriptions are incompatible with the Maxim of Quantity.

In the current studies we focus primarily on over-descriptions, because evidence suggests they are more common than under-descriptions (Deutsch & Pechmann, 1982; Ferreira, Slevc, & Rogers, 2005; Olson, 1970; Sedivy, 2003). In addition, the assumption that over-descriptions are avoided has played a critical role in debates about the architecture of language processing. These discussions have centered on restrictive post-nominal modifiers, including prepositional phrases. A sentence such as *Mary put the frog on the towel into the box* contains a temporary syntactic ambiguity, because *on the towel* could be either a modifier of *frog* (as it turns out to be, in this case), or the location to which the frog is to be moved. The Garden-Path model of parsing assumes that syntactic representations are created serially and that the system's initial preference is for syntactic simplicity (Frazier, 1978). Moreover, the parser is an informationally encapsulated module, consulting only syntactic principles during the creation of the single initial parse. If the initial parse is not viable (as is the case in the above example), then the more complex structure will be built or reactivated. Thus, according to the Garden-Path model, listeners' preference to interpret the phrase *on the towel* as a location rather than a modifier is a consequence of the parser's preference to avoid postulating potentially unnecessary nodes.

In contrast, other models assume that the system considers all alternatives in parallel when there is an ambiguity, and that the parser consults all potentially relevant sources of information (MacDonald, Pearlmutter, & Seidenberg, 1994; Spivey & Tanenhaus, 1998; Tanenhaus & Trueswell, 1995). If the sentence occurs in a linguistic or visual context in which there is only one frog, the listener will apply his knowledge of Gricean principles and infer that *on the towel* is not a modifier, since it is not necessary. If there is more than one frog, the putatively more complex structure will be favored from the outset, because the modifier interpretation allows the semantic/referential system to uniquely identify the appropriate referent. Thus, the assumption that listeners assume a contrastive interpretation when they hear a complex description is critical in explaining comprehenders' tendency to experience a Garden-Path for structures such as *put the frog on the towel into the box* when only one frog exists. (Other Garden-Path forms that turn on an ambiguity between a modifier and non-modifier interpretation are explained in a similar way). If people do not have this expectation, then the tendency to prefer the non-modifier interpretation must have a different cause (e.g., the preference for syntactic simplicity).

One early model that assumed this application of the Maxim of Quantity (i.e. the principle of parsimony) to the parsing of syntactically ambiguous sentences is Referential Theory (Altmann & Steedman, 1988; Crain & Steedman, 1985). Crain and Steedman investigated structures such as the reduced relative (e.g., *The neighbor warned about the criminal called the police*), which have long been known to be difficult to understand (Bever, 1970). The Garden-Path model attributes this difficulty to the parser's preference for the syntactically simple, main clause interpretation of *warned about the criminal*. But according to the Referential Model, the active structure is built initially because the sentence occurs in a null discourse context and there is therefore no referential support for the complex reduced relative form. Thus, if the reduced relative occurred in a context that established the existence of a set of neighbors, then the sequence *warned about the criminal* would be interpreted as a modifier right away, because the comprehender needs modification to identify the correct referent of *neighbor*. In an offline judgment task, Crain and Steedman observed large effects of context on decision times for the reduced relative, consistent with their hypotheses.

These results were later challenged in a series of experiments with the same logic but employing eye-movement monitoring to allow measurement of online processing (Ferreira & Clifton, 1986). Reduced relative sentences and sentences with temporarily-ambiguous prepositional phrases (e.g., *Mary put the book on the table into her bag*) were placed in contexts like the ones created by Crain and Steedman (1985). Ferreira and Clifton found that performance on a question–answering task was consistent with what Crain and Steedman had observed, but online measures of processing (e.g., first-pass reading times) showed that the parser initially preferred the syntactically simpler structure. They concluded that the referential principle applied during processing, but only on the output of the parser. The parser initially selected the simpler structure, but when reanalysis was forced syntactically, contextual information helped the comprehension system identify the correct structure. (These issues were considered in further elaborations of the Referential Model; Altmann & Steedman, 1988; Clifton & Ferreira, 1989; Trueswell, Tanenhaus, & Garnsey, 1994.)

The debate has been renewed with the introduction of a new and powerful technique for studying language processing, the Visual World Paradigm. In this paradigm, spoken language comprehension is assessed by monitoring eye movements as people view visual displays (either real-world or displayed on computer monitor). Cooper (1974) demonstrated that people reliably fixate pictures that are related to concurrent spoken language, and that eye movements are systematically time-locked to related spoken words. During the past decade, this paradigm has become popular for studying the processing of temporary syntactic ambiguities, particularly those involving ambiguous prepositional phrases (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell, Sekerina, Hill, & Logrip, 1999). Tanenhaus et al. (1995) investigated this type of ambiguity in an experimental task in which participants were required to move objects in response to spoken instructions.

The displays contained either one or two relevant objects to be moved (in Figs. 1A and B, one apple or two apples), and instructions to move the objects were either syntactically ambiguous (1) or unambiguous (2).

(1) *Put the apple on the towel in the box*. (ambiguous)
(2) *Put the apple that's on the towel in the box*. (unambiguous)

When the visual context contained just one apple and the instruction was ambiguous, participants fixated the incorrect location (the empty towel) shortly after hearing the prepositional phrase *on the towel* on 55% of trials. With the unambiguous instruction, participants almost never looked to the incorrect location. When the visual context contained two apples, the empty towel was fixated much less often overall, and there was no difference between ambiguous and unambiguous instructions (15 and 12% of trials had fixations on the empty towel, respectively). This pattern is consistent with the predictions of the Referential Model, which assumes that listeners expect speakers to follow the Maxim of Quantity. If the relevant visual world contains only one apple, a modifier is unnecessary and the ambiguous prepositional phrase is therefore initially interpreted as a location, which results in looks to the empty towel. In contrast, with two apples, a modifier is required to allow the correct referent to be identified, and the ambiguous prepositional phrase is therefore immediately interpreted as the contextually-necessary modifier.

These findings support the idea that over-descriptions are problematic for listeners, but the evidence is indirect. To directly test the Gricean Maxim of Quantity,

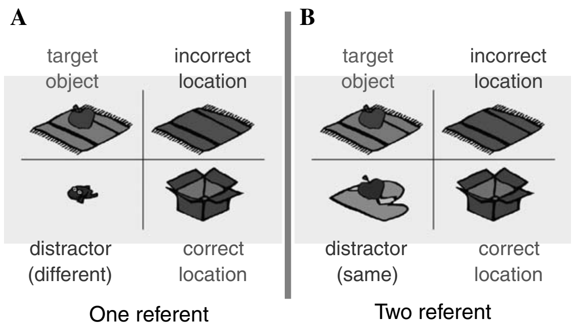

Fig. 1. (A and B) One referent (apple) and two referent (two apples) displays used in Tanenhaus et al. (1995) and Spivey et al. (2002).

we must examine whether speakers avoid producing over-descriptions, and we must assess to what extent listeners are misled by them. We did this in the experiments reported below. First, a production study was conducted in order to determine the types of utterances that participants generated when instructing another person to move the target object to either of the empty locations (see Figs. 2A and B). In the second experiment, participants rated the acceptability of four different instructions (see Table 1).

In the third experiment, we monitored the eye movements of participants as they moved objects in response to those same instructions. If the Gricean Maxim of Quantity is used in the formulation of referential expressions, then speakers should not over-describe; for example, they should rarely refer to an apple in a display such as Fig. 2A as *the apple on the towel*. Listeners should judge such an utterance/visual world combination to be inappropriate or unacceptable, because the modifier is informationally unmotivated. Listeners' eye movements

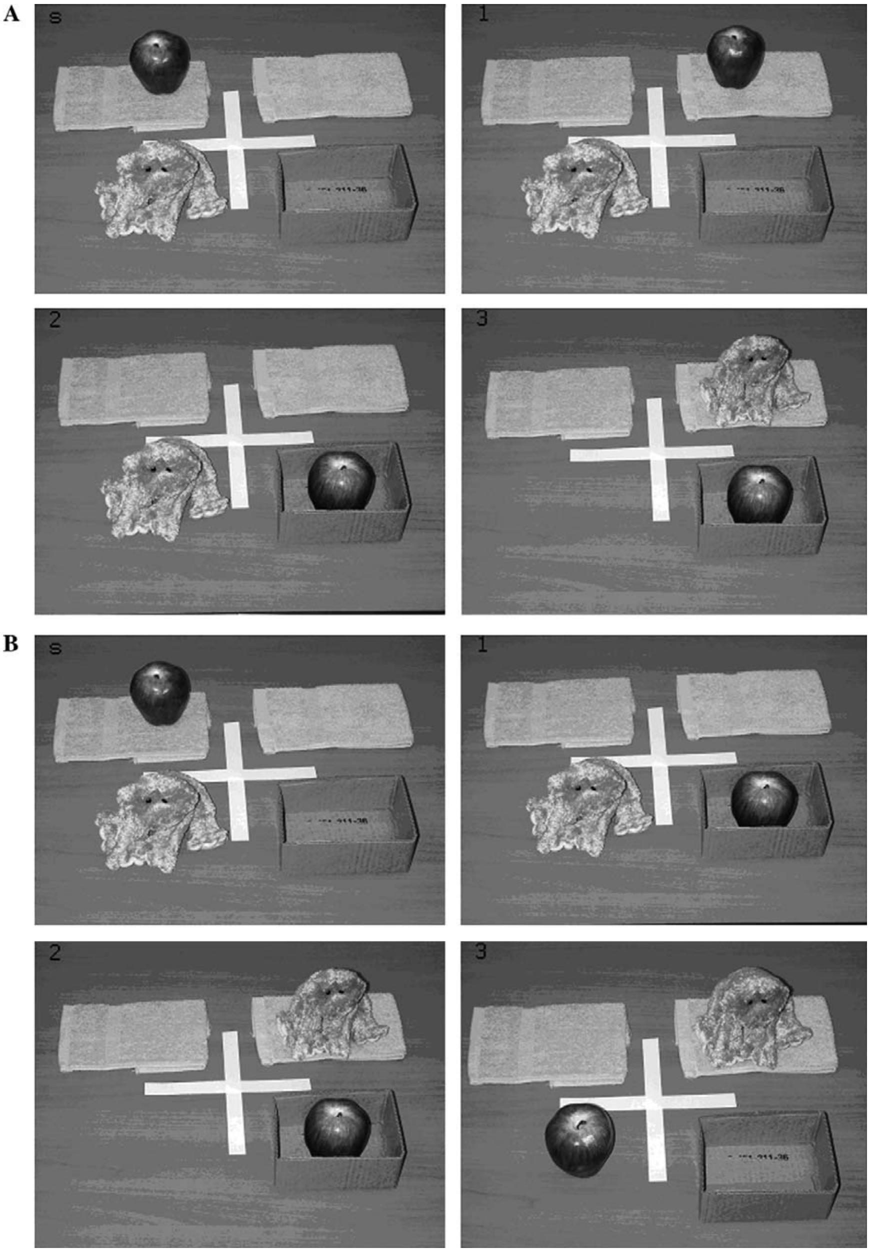

Fig. 2. (A) Example stimulus for Experiment 1 (matching condition). (B) Example stimulus for Experiment 1 (different condition).

Table 1
Instructions tested in Experiments 2 and 3

	Instruction type	
	Prepositional phrase modifier	Location type
(3) Put the apple on the towel	No	Matching
(4) Put the apple in the box	No	Different
(5) Put the apple on the towel on the other towel	Yes	Matching
(6) Put the apple on the towel in the box	Yes	Different

should reflect their tendency to think that the ambiguous prepositional phrase is a location, and there should be frequent looks to the empty towel. If the Gricean Maxim of Quantity is not used by speakers and listeners or if other principles come into play and override it, then we would expect the following: (1) Speakers should sometimes over-describe, (2) listeners should judge over-described sentences as acceptable, and (3) eye movements should show little evidence that listeners interpret the empty towel as a location, because listeners would not necessarily expect modifiers to be used to implicate the existence of a set.

These three experiments together allow us to determine whether speakers avoid over-descriptions, whether listeners find them infelicitous, and whether listeners perform less efficiently when they encounter them. Alternatively, we might find dissociations between production performance, explicit judgments of sentences, and implicit comprehension performance.

Experiment 1: Production

In Experiment 1, we examined the types of utterances that participants produced in contexts such as those shown in Figs. 1A and B. Our two referent context (Fig. 1B) was somewhat different, however, in that only one of the potential target objects was on a location object (for example, one apple was on a towel but the second apple was by itself). Participants instructed a confederate to move real objects to both "matching" and "different" locations. In both conditions, this object to be moved was always on or in some other object (e.g., an apple on a towel). In the "matching" location condition, the instruction was to move the target object from one location to another of the same type (e.g., the apple on a towel was to be moved to another towel). In the "different" location condition, the target object was to be moved to a different type of location (e.g., the apple on a towel was to be moved to a box). Participants were shown four pictures (see Figs. 2A and B) and were required to give three instructions to another "subject" (in reality, a confederate) corresponding to the move shown between each successive pair of pictures. Participants were told prior to the experiment that the "other participant" would not necessarily have the objects in the same configuration as the one shown in the pictures. This information was provided in order to prevent subjects from using directional terms (i.e., "move the apple diagonally" or "move the apple to the right") in their instructions. We constrained the instructions in this way because our experiments attempt to determine how much modification is usually provided and how such modifiers are interpreted. It was therefore critical that participants use linguistic terms that did not depend on objects being arranged in a 2×2 configuration.

The main purpose of this study was to determine what type of instruction a naïve participant would give to get a person to move a target object to one of the other available locations. In particular, we asked whether the speaker would produce over-descriptions such as *put the apple on the towel in the box*. In addition, we were interested in whether participants in the matching location condition would provide a modifier for the location because the object to be moved was already on an object of that same type. Thus, if the apple was on a towel and the new location was a different towel, they might tend to say *put the apple on the other towel*.

Methods

Participants

Ten students from the undergraduate subject pool at Michigan State University participated in the study for course credit. Participants were native speakers of English and had normal or corrected-to-normal vision.

Materials

Stimuli for this experiment consisted of four pictures printed on a single sheet of paper. Pictures were labeled in the upper left corner as S, 1, 2, and 3, as shown in Figs. 2A and B. The images were printed in grayscale to prevent participants from using color to indicate which object had to be moved. The difference between each pair of pictures (i.e., S–1; 1–2; 2–3) consisted of a single movement of one object. We created a total of 53 visual displays: five were for practice, 24 were experimental items, and 24 were filler displays. In experimental trials, the target and distractor objects were always on the left side of the display and possible location objects were on the right (as in Spivey et al., 2002). In the practice and filler trials, target, distractor, and location objects all had an equal probability of appearing in any of the four quadrants of the display, and no constraints were placed on direction of movement. Practice trials consisted of three two-referent trials and two one-referent trials. Filler trials consisted of 16 one-referent trials and eight two-referent trials. Eight of the one-referent filler displays had no duplicated location objects (for example: an apple on a towel, a frog, an empty box, and an empty bowl). The other eight one-referent filler displays were like the one shown in Fig. 2A. The two-referent filler displays had two identical target objects (e.g., two apples); one of these objects was located on another object and the other was by itself. The other two quadrants of the display were occupied by two distinct location objects so that no location objects were duplicated in the display.

Procedure

The confederate was placed in the waiting area of the laboratory and waited for the participant to arrive. Once he or she did, the experimenter addressed both in a similar manner (for example, he always asked each of them to give their names). The confederate and the participant both read the instructions and completed consent forms. They were then shown all of the objects that would be used in the experiment and were told their category label (e.g., apple, towel, and box; see, Table 2).

The participant and confederate were then seated on opposite sides of a table facing one another. The experimental stimuli and a microphone were placed in front of the participant, and the objects corresponding to the stimuli were placed in front of the confederate. For each trial, the participant gave three instructions, pausing after each to allow the confederate to move an object. The confederate moved the real object corresponding to each instruction and responded with "okay" after the move was completed. The experimenter added and removed objects between trials. A divider was placed between the participant and confederate after the third practice trial to ensure that the participant would not use physical cues such as eye gaze or hand location in generating utterances (e.g., *Now move that apple over there*). Participants were not given feedback concerning the quality of their instructions nor were they corrected if an instruction was referentially ambiguous. Each participant completed five practice trials followed by 24 critical trials (six trials in each condition) and 24 filler trials. The order of critical and filler trials was randomly assigned for each participant. The utterances produced by each participant were recorded on audiotape, and only the first instruction for each critical trial was transcribed. After completing all trials, the participant was debriefed. The entire session lasted about 45 min. Participants were tested individually.

Table 2
Objects used in Experiments 1–3

Target & distracter objects	"in" Locations	"on" Locations
Apple	Basket	Felt
Banana	Bowl	Glove
Bat	Box	Napkin
Brush	Flowerpot	Sock
Candle	Mug	Sponge
Car	Pan	Towel
Dinosaur		
Football		
Frog		
Jeep		
Lizard		
Plane		

Design

The experiment had a 2×2 within-participants design: Visual contexts contained either one or two instances of the object to be moved (one-referent versus two-referent condition), and the location type to which the object was to be moved either matched (towel to towel) or was different (towel to box). Analyses were conducted with both participants ($F1$) and items ($F2$) as random effects.

Responses were categorized as follows: If the target object (e.g., apple) was not modified, the utterance was labeled "bare target". This categorization applied even if the location was modified by a word such as *other*. If the target object was modified, the utterance was labeled "modified target". This response category does not distinguish between reduced and non-reduced descriptions—i.e., *put the apple on the towel in the box* and *put the apple that's on the towel in the box* were both categorized as modified target. The number of modified target responses was divided by the total number of utterances produced for each condition. This proportion was the dependent variable for statistical analyses.

Orthogonal to this categorization scheme, we also tabulated whether a pre-nominal modifier such as other was used to describe the location to which the object was to be moved. A response either included other (or opposite, or another similar term) or it did not. In this scheme, we ignored whether the noun phrase referring to the target object was modified or unmodified. This analysis examined whether speaker would likely produce the instruction put the apple on the towel when an object is moved from one location to another location of the same type.

Results

One utterance was eliminated from the analysis because the participant did not give a correct instruction for the move shown.

Results are shown graphically in Fig. 3, where the proportion of modified target utterances is broken down by visual context (one vs. two-referent) and location type (matching or different). Floating error bars show the 95% confidence interval for the pairwise comparison between adjacent condition means for the participants data (Masson & Loftus, 2003). Let us consider first under what conditions people produced modified target utterances. These should occur in the two-referent condition but not in the one-referent condition, because modified target utterances in the one-referent condition would constitute over-descriptions. Consistent with the Gricean Maxim of Quantity, we did indeed find that modified target utterances were more common in the two-referent condition than in the one-referent condition (98% vs. 30%), $F1(1,9) = 108.08$, $p < .001$, $F2(1,5) = 257.83$, $p < .001$, $\min F'(1,14) = 76.15$, $p < .001$. Of course, this is not surprising, as complex noun phrases were essential for effective communication when two objects of the same type were present. Less compatible with the Maxim of Quantity is our finding that modified target instructions occurred on 30% of trials overall in the one referent condition. A one sample *t*-test revealed that the number of overdescriptions produced in the one-referent condition was significantly greater than 0, $F1(1,9) = 11.0$, $p < .01$, $F2(1,5) = 102.96$, $p < .001$, $\min F'(1,11) = 9.94$, $p < .01$. (Note for this analysis the *t* values were changed to *F*'s to calculate min*F*′.) These over-descriptions occurred twice as often when the location to which the object was to be moved was matching rather than different (40% vs. 20%), but this difference was not significant, being smaller than the 95% confidence interval of .30 for the comparison between condition means. There was no interaction between number of referents and matching/different location, $F1(1,9) = 3.56$, $p = .10$, $F2(1,5) = 3.01$, $p = .15$, $\min F'(1,12) = 76.15$, $p = .23$.

The important finding from this analysis is that over-descriptions were surprisingly common in the one referent display; overall, almost a third of utterances included the unnecessary modifier. Again, recall that we refer to the modifier as unnecessary because the one-referent displays contained just one apple (for example), so the expression *the apple* was sufficient to allow referent identification. Of course, the modifier might be necessary or at least compelling to the speaker for some other reason; we will return to this possibility in the General Discussion.

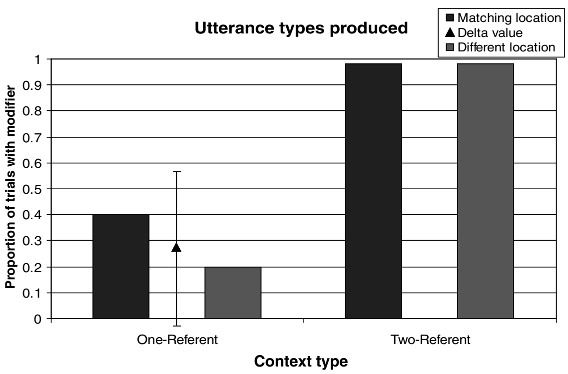

Fig. 3. Results from Experiment 1, floating error bars show the 95% confidence interval for the pairwise comparison between adjacent condition means.

Now, let us consider whether people simply said *put the apple ... on the towel* when the apple was to be moved from one towel to another (i.e., the matching condition). The answer is that they did not. Regardless of whether there was one referent or two referents in the display (97% vs. 96%; both $F\text{s} < 1$), people almost always included a modifier when the location was matching (96.5%), and almost never included one when the location was different (4% of trials), $F1(1,9) = 644.78$, $p < .001$, $F2(1,5) = 198.24$, $p < .001$, $\min F'(1,8) = 152.62$, $p < .001$ (see Table 3 for a breakdown of the types of modifiers produced in the matching condition). The interaction between these two variables was not significant. Thus, it appears that if a speaker intended to make a listener move an apple from one towel to another, he or she would modify the location object so the resulting utterance would most likely be *put the apple on the other towel*. In other words, speakers did not under-describe.

Discussion

These production data have yielded some unexpected findings from the point of view of evaluating the Gricean Maxim of Quantity. Let's begin with results that were expected. First, consistent with the principle which states that speakers should provide as much information as necessary, over 95% of the time speakers modified a noun when there was more than one of the object that it denoted. Thus, speakers avoided under-descriptions. However, they also included the modifier one-third of the time when it was not necessary (i.e., in the one referent display conditions). Thus, it appears that speakers produce a surprising number of over-descriptions—far more than would be

Table 3
Modifiers produced in matching condition in Experiment 1

	One-referent	Two-referent
(1) Other (i.e. the other towel) (%)	72	58
(2) Opposite (i.e. the opposite towel) (%)	6	10
(3) Directional (i.e. the lower-right towel) (%)	13	18
(4) Post-nominal (i.e. the towel on the right) (%)	6	10

expected if the Gricean Maxim of Quantity determined the form and content of people's utterances. In addition, speakers almost always included the word *other* (or *opposite* or some semantic equivalent) in the matching location (towel-to-towel) condition. It was clearly pragmatically relevant to them that the apple was to be moved from one towel to another, and so this was invariably marked in speakers' utterances with the modifier *other* or *opposite*.

In summary, speakers are fairly likely to over-describe objects, but they avoid under-descriptions. Target objects were almost always described with a modifier when there were two of the same type, and a destination was referred to as *other* or *opposite* when it was of the same type as the current location. In the next experiment, we assess whether listeners judge under- and over-described utterances to be infelicitous, as they should according to the Maxim of Quantity.

Experiment 2: Listener judgments

The purpose of this experiment was to assess whether people find over- and under-described utterances infelicitous. Subjects judged spoken utterances by examining two photographs presented side-by-side on a computer screen while they listened to a spoken instruction over headphones. The pictures were the same as those used in Experiment 1 (see Figs. 2A and B, pictures *S* and 1). In the one referent condition, there was (for example) an apple on a towel, a frog, an empty towel, and an empty box in the first picture. The pictures differed only in that the relevant object had been moved either to the matching location (towel to towel), or to a different location (towel to box) in the second picture. In the two-referent condition, there were two relevant objects (e.g., two apples), an empty towel, and an empty box in the first picture. Again, the apple on the towel was moved either to a matching or to a different location in the second picture. The participants' task was to judge whether the instruction given was a good one to bring about the change between the first and second pictures. Participants rated the instructions on a scale of 1 to 5, where 1 was an incorrect instruction, 3 was an adequate instruction, and 5 was a perfect instruction.

The instructions people received are shown in utterances (3–6) from Table 1, reprinted below:

(3) *Put the apple on the towel.* (matching)
(4) *Put the apple in the box.* (different)
(5) *Put the apple on the towel on the other towel.* (matching)
(6) *Put the apple on the towel in the box.* (different)

Instructions (3) and (4) correspond to the "bare target" responses from Experiment 1, and instructions (5) and (6) correspond to "modified target" responses. However, one distinction that should be noted is that Experiment 2 examined temporarily ambiguous modified target instructions, whereas the utterances produced in Experiment 1 were not necessarily ambiguous. The bare target instructions should be judged as "bad" in the two-referent conditions, regardless of whether the location is matching or different, because they are under-descriptions of the sort that do not permit unambiguous identification of the intended referent. In addition, instructions involving a move from one towel to another should also be judged bad if the location object is not modified, for example with the word *other* or *opposite*, because, based on our production results, such an instruction is also an under-description (even though it ultimately does not produce any type of pragmatic confusion, since an object must be placed in a different location to be correctly described as having been "moved"). The two modified target instructions should be rated high in the two-referent condition but low in the one-referent condition, because modified instructions in the former condition are over-descriptions and therefore violate the Gricean Maxim of Quantity. Again, whether the location is matching or different should not matter, because the location is not relevant; all that matters is that with only one apple in the display, it is not necessary to describe it as being on a towel.

Method

Participants

Twenty-two people drawn from the same pool as for Experiment 1 participated in this experiment. All were native speakers of English, had normal or corrected-to-normal vision, and none had participated in Experiment 1.

Materials

The stimuli consisted of "before" and "after" pictures presented side by side on a computer screen. The pictures were the same as pictures S and 1 from the previous experiment, except they were shown in color. Each participant completed 10 practice trials and 72 regular session trials. The latter consisted of 24 experimental (six in each condition, as in the previous experiment) and 48 filler trials. The practice, filler, and experimental trials were designed so that half of the instructions were "good" and half of the instructions were "bad". A "good" instruction was one that unambiguously named the target object and goal, and the instruction was correct for the movement shown in the pictures. "Bad" instructions used either incorrect category labels (target or location) or incorrect modification (for example, *upper-left apple* when the upper-right apple was moved). Eighteen of the filler trials were "good" and thirty were "bad", as all of the experimental instructions were "good" except for the bare target instructions in the two-referent condition. Sound files were recorded by a female native speaker of English, digitized using Computerized Speech Laboratory 4400 (Kay Elemetrics), and converted to .wav format.

Apparatus

The experiment was programmed using Superlab experimental software (Cedrus) and auditory stimuli were played through Sony MDR-CD60 headphones. Participants completed the experiment on a Dell Optiplex GX 400 computer.

Design and procedure

The experiment had a $2 \times 2 \times 2$ design. Visual contexts contained either one or two possible target referents, the location type either matched the object that the target was currently on or was different, and the instruction type either contained a prepositional phrase modifier or did not (modified target vs. bare target). All variables were manipulated within-subjects, but display type was a between-items variable (the other two variables were within-item).

The bare target instructions in the different condition were created by digitally removing the modifier (the first prepositional phrase) from the modified target instructions. For the bare target instructions in the matching condition, a separate utterance like example (7) was created, and then the first prepositional phrase was digitally removed. The creation of a separate utterance ensured that the two bare target instructions had similar prosody. The resulting utterances sounded as natural as the unaltered utterances (speech files can be obtained at http://eyelab.msu.edu/pubs/ebf/).

(7) *Put the apple in the box on the towel.*

Before the experiment began, participants were shown all of the objects that were used and were told what label would be used for each (see, Table 2). Participants then read the instructions and signed a consent form, after which the experimenter asked if participants had any questions. Participants were instructed to rate the spoken instruction on a scale of 1–5, where a 1 was an incorrect instruction (i.e., the wrong category label or incorrect modification), a 3 indicated an adequate instruction, and a 5 should be given to an instruction that could not be better. Each trial began with a message on the screen reading, "Press any key when ready", which was followed by a 1000 ms interstimulus interval. The visual display was shown for 2000 ms before the utterance was played out through headphones. The display remained on the screen until the participant responded. The order of trials was randomly assigned for each participant. Participants were tested individually, and the entire experimental session lasted ≈25 min.

Results and discussion

Results are presented in Figs. 4A and B. Floating error bars show the 95% confidence interval for the pairwise comparison between adjacent condition means for the participants data (Masson & Loftus, 2003). The result of a $2 \times 2 \times 2$ repeated measures ANOVA revealed a significant interaction, $F1(1,21) = 48.44$, $p < .001$, $F2(1,5) = 42.25$, $p < .01$, $\min F'(1,12) = 22.56$, $p < .001$. The following analyses are broken down by display type because the Gricean Maxim of Quantity makes distinct predictions for each type of context. The inferential statistics for the main effects and interactions are presented in Table 4 (n.s. indicates not significant).

One-referent context

The main effects of location type and instruction type were both significant. The different location conditions produced higher ratings than the matching location conditions (4.61 vs. 3.93), and the modified target instructions were rated as better than the bare target instructions (4.52 vs. 4.02). In addition, the location $\times$ instruction interaction was significant (see Fig. 4A). Paired comparisons revealed four significant effects. First, the difference between the location conditions within the bare target instructions exceeded the 95% confidence interval of .31 for the difference between condition means. The different condition was rated higher than the matching condition (4.91 vs. 3.12). This result is not surprising considering the disparity that was observed in the production data (recall that the unmodified instruction was produced on 80% of trials in the different condition and 60% of trials in the matching condition). This result demonstrates that people explicitly judge under-descriptions of the location to be less felicitous than appropriate descriptions. It appears that when there are two identical locations, a modifier (either pre- or post-nominal) is obligatory. The difference between the two location conditions with the modified target instructions also exceeded the 95% confidence interval of .30 for the difference between condition means. The matching location was rated higher than the different location (4.74 vs. 4.3). Again, this result was predicted because the production data showed that speakers modified the target object on 40% of trials in the matching condition. In addition, speakers also modified the location object on 97% of trials in the one-referent matching condition. It is therefore not surprising that comprehenders rated the

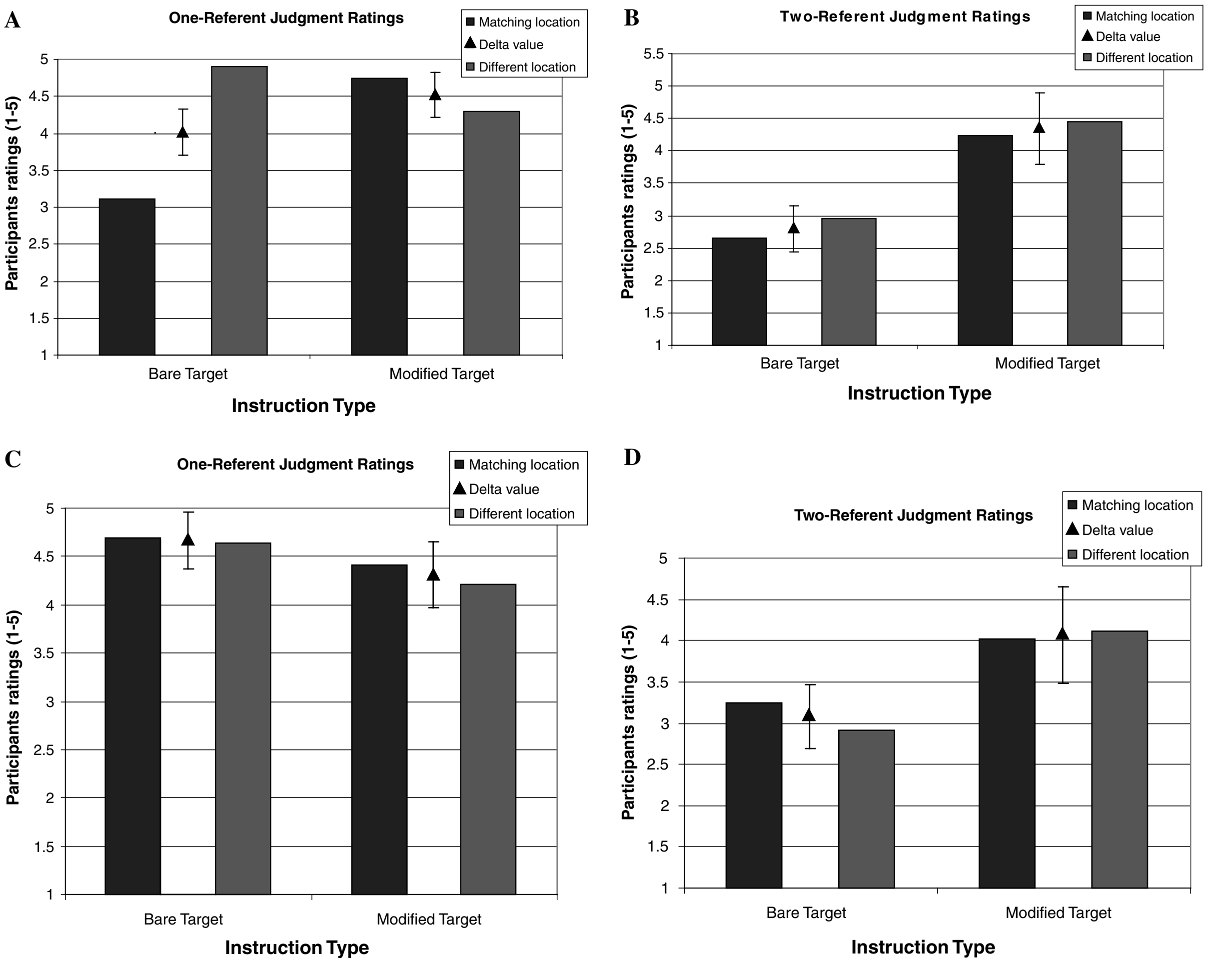

Fig. 4. (A) Results from Experiment 2, one-referent condition, floating error bars show the 95% confidence interval for the pairwise comparison between adjacent condition means. (B) Results from Experiment 2, two-referent condition, floating error bars show the 95% confidence interval for the pairwise comparison between adjacent condition means. (C) Results from Experiment 2b, one-referent condition, floating error bars show the 95% confidence interval for the pairwise comparison between adjacent condition means. (D) Results from Experiment 2b, two-referent condition, floating error bars show the 95% confidence interval for the pairwise comparison between adjacent condition means.

modified target instruction higher, as it contained both types of modification. Furthermore, the difference between the two instruction conditions in matching location displays exceeded the 95% confidence interval of .27 for the comparison between condition means. The bare target instruction was rated as worse than the modified target instruction (3.12 vs. 4.74).

The difference between two different location instructions also exceeded the 95% confidence interval of .38 for the comparison between condition means. However, for this comparison the bare target instruction was rated higher than the modified target instruction (4.91 vs. 4.3), which is consistent with the Maxim of Quantity. We also observed a weak preference for utterances that modified the target object inappropriately (i.e., in the one-referent condition) when compared to a display that required modification of the target object.

Two-referent context

There was a main effect of instruction type (see Fig. 4B). The modified target instructions were rated significantly higher than the bare target instructions (4.35 vs. 2.8). This result is not surprising; under-descriptions can compromise communication, as an under-described utterance will not permit a listener to identify the correct referent from a set. A similar pattern was observed for both the matching location and the different location. The location type × instruction type interaction in the two-referent context was not significant. Ratings in the

Table 4
Analyses of variance for Experiment 2 by subjects, items, and minF

Display type	One-referent	Two-referent
Location type		
By subjects	$F1(1,21) = 48.93$, $p < .001$	n.s.
By items	$F2(1,5) = 107.67$, $p < .001$	n.s.
MinF'	$F3(1,24) = 33.64$, $p < .001$	n.s.
Instruction type		
By subjects	$F1(1,21) = 11.25$, $p < .01$	$F1(1,21) = 72.59$, $p < .001$
By items	$F2(1,5) = 25.28$, $p < .01$	$F2(1,5) = 74.42$, $p < .001$
MinF'	$F3'(1,24) = 7.79$, $p < .05$	$F3(1,16) = 36.74$, $p < .001$
Location × instruction		
By subjects	$F1(1,21) = 98.97$, $p < .001$	n.s.
By items	$F2(1,5) = 159.47$, $p < .001$	n.s.
MinF'	$F3(1,21) = 61.07$, $p < .001$	n.s.

two-referent condition showed a similar pattern to the production data from Experiment 1, and indicate that participants were sensitive to contexts containing a referential ambiguity with respect to the target object.

What was unexpected was the finding that the under-description in the one-referent context which involved failing to distinguish one location from another when they were of the same type (e.g., when the situation required moving an object from one towel to another) produced a similar result as the under-descriptions in the two-referent context in which there were two target referents. Again, the finding that modified target utterances in the one-referent condition produced higher ratings than the bare target instructions appears to be inconsistent with the Maxim of Quantity. However, this effect is primarily driven by the low rating given to the bare target instruction in the matching condition. Therefore, our initial conclusion that the modified target instructions were preferred over bare target instructions is compromised by the fact that the two bare target instructions are not equivalent because the matching condition contains an under-description of the location and the different condition does not. We therefore ran a follow-up experiment using the same materials and design, but we replaced utterances like (3) *put the apple on the towel* with utterances such as *put the apple on the other towel*. The results from this follow up study revealed that the bare target instruction in the matching condition produced a similar rating as the bare target instruction in the different condition (4.69 vs. 4.64), and both were rated higher than the modified target instructions (4.41 vs. 4.21) (see Figs. 4C and D). However, this difference was not statistically significant.The only significant effect in the follow-up experiment was a main effect of instruction type with the two-referent context $F1(1,21) = 23.73$, $p < .01$, $F2(1,5) = 12.33$, $p < .05$, min$F'(1,11) = 8.11$, $p < .05$.

Over-descriptions, then, do not seem to be considered infelicitous in one-referent contexts, as they are given essentially equivalent ratings as concise instructions, a finding that is inconsistent with the Gricean Maxim of Quantity. Before attempting to explain this result, it is important to determine whether it is consistent with implicit performance in a comprehension task. We turn to this question in the next experiment.

Experiment 3: Eye movements in the Visual World Paradigm

In this experiment, we recorded participants' eye-movements as they moved objects in response to instructions, using a procedure similar to that described by Spivey et al. (2002). Two variables were manipulated: the location to which the target was to be moved was either matching or different, and the instruction either contained a modifying prepositional phrase or did not. Visual context (one vs. two-referent) was not manipulated in this experiment; critical trials tested only the one referent displays. The instructions were the same as in Experiment 2 (see, Table 1). We made three general predictions based on the results of Experiments 1 and 2. First, performance should differ in the two bare target instructions (Table 1, utterances [3–4]). Specifically, participants' eye movements when they hear (3) should indicate mild confusion. We make this prediction because we observed that speakers would never produce an instruction like (3) in such a context, and also because we found that listeners judge it to be infelicitous relative to its context.

Second, by comparing (3) and (5), we can measure the effect of having an underspecified location object (3) with an instruction that has an overspecified target but an appropriately modified location, because in (5), the location contains the modifier *other*.

Third, we can compare performance between instructions (4) and (6) to evaluate the cost, if any, of over-describing. Instruction (4) is appropriately concise in the one referent display, but instruction (6) contains an unnecessary modifier. The first experiment, however, showed that speakers often produce such an instruction (about one-third of the time), and the second experiment revealed that listeners judge such over-descriptions to be no worse than concise descriptions. Given these two findings, we predicted no difference in eye-movement behavior for (4) versus (6). On the other hand, the results of Spivey et al. (2002), predict that we should observe eye movements to the empty towel when listeners hear the unnecessary prepositional phrase. Recall that these looks to the empty towel are taken to indicate that listeners interpret the prepositional phrase (at least momentarily) as a potential location—in other words, that they are Garden-Pathed (Spivey et al., 2002). The results of Experiment 1 and 2 do not lead us to make the same prediction because our production data showed that speakers pre-nominally modify the location object in the one referent, matching location condition over 90% of the time. (This result may be inflated because of our manipulation to discourage directional modification, see Experiment 1). Therefore, if a listener hears an instruction like *put the apple on the towel*, then he or she should immediately interpret *on the towel* as a modifier because it does not contain the word *other* or *opposite*. This prediction is further supported by the finding that speakers often produce unnecessary modifiers. So why might comprehenders immediately fixate the empty towel when they hear *put the apple on the towel*? If it is true that listeners never encounter this type of instruction in this sort of context, then either production frequencies and comprehension tendencies are sometimes inconsistent with one another, or saccades to the empty towel reflect something other than Garden-Pathing.

Method

Participants

Twenty-one students from the undergraduate subject pool at Michigan State University participated in this study for course credit. None had participated in the previous two experiments, all were native speakers of English, and all had normal or corrected-to-normal vision.

Materials

The utterances were pre-recorded by a female native speaker of English, digitized using Computerized Speech Laboratory 4400 (Kay Elemetrics), and then converted to .wav format. Each trial consisted of three instructions, as in the relevant previous studies (Spivey et al., 2002). In experimental trials, the target and distractor objects were always on the left side of the display and possible location objects were on the right. Filler trials were created so that targets, distracters, and location objects all had an equal probability of appearing in any of the four quadrants of the display, and no constraints were placed on direction of movement. The critical instruction (which were the same sound files used in Experiment 2) was always the first one, and the subsequent two instructions were fillers. The second and third instructions for both experimental and filler trials were unambiguous bare target instructions. Filler displays consisted of six one-referent trials and six two-referent trials. Three of the one-referent filler displays had no duplicated location objects (for example, an apple on a towel, a frog, an empty box, and an empty bowl). The first sentence for each of these trials had bare target instructions and none was ambiguous. The other three one-referent filler displays were like the one shown in Fig. 1A. The first instruction for each of these trials had post-nominally modified location objects such as *Put the apple on the towel on the right*. The two-referent filler displays were similar to the one shown in Fig. 1B except that the second potential target object was not located on a goal object. Half of these trials had modified target instructions like examples (5) and (6). The other half used pre-nominal (directional) modification of the target object to distinguish between the two identical referents (for example, *put the lower right apple in the box*). Each participant was run through 12 experimental and 12 filler trials. Two lists were created so that for each experimental display, there were both bare target and modified target instructions. Each visual display only appeared once per list and lists were counterbalanced so that half of the instructions were bare target and half were modified target. Half of the participants received one list and the other half received the other.

Apparatus

We recorded eye movements using an ISCAN model ETL-500 head-mounted eyetracker (ISCAN). Subjects were able to view 103° of visual angle horizontally and 64° vertically. The ETL-500 consists of two cameras, both attached to a visor. The left eye was sampled at 60 Hz. Both video streams were recorded using a Sony Digital 8 video recorder. The merged video was digitized at 30 Hz using a Sony Mini-DV recorder. The pre-recorded audio files were played to participants using Superlab Pro (Cedrus). Hand coding of the data was performed from the onset of the critical instruction to the time at which the participant let go of the target object.

Design and procedure

We manipulated two variables. First, the location to which the object was to be moved was either matching or different, and second, the instruction was either a bare target or a modified target. Prior to the experiment, participants were shown all of the objects that would be used and told the names of each. The eye-tracker was then fitted on the participant and calibrated. After calibration, one practice trial was completed to familiarize the participant with the procedure. Participants were allowed to view each display as it was being set up, and so participants had several seconds in which to view each display before the instructions began (as in previous studies; Spivey et al., 2002; Tanenhaus et al., 1995). Participants performed the task while standing up and objects were located on a table at a comfortable height in front of them. Participants were tested individually, and sessions lasted ≈30 min.

Results and discussion

One participant was excluded from the analyses because his/her eye tracking data could not be coded reliably. For all remaining participants, data were analyzed in segments corresponding to words within the utterance and eight 300 ms time windows from the offset of the utterance. Word time windows were identified by aligning utterances on a word-by-word and trial-by-trial basis, as recommended by Altmann and Kamide (2004). Proportions corresponding to the probability of fixating a region were calculated by determining the number of trials that contained a fixation in one of the four quadrants of the display during either a word or a 300 ms window. The number of trials with a fixation was divided by the total number of trials per condition. Thus, the dependent measure was the proportion of trials with a fixation to a region during one of the time windows. Statistical analyses consisted of one-way ANOVAs conducted on arcsine transformed proportions for each word and 300 ms time window. The results are shown in Table 5. The 95% confidence intervals for the difference between means in each time region (labeled with a delta symbol), calculated from the analysis by participants, are given in Figs. 5–7. Values of the confidence intervals are back-transformed from arcsine-transformed proportions that served as the dependent measure in the analyses.

Effects of under-describing the location: "*Put the apple on the towel*" *versus* "*Put the apple in the box*". Fig. 5A shows looks to the target object (e.g., the apple), and Fig. 5B shows looks to the correct location (e.g., the box in the different condition and the towel in the matching condition). In both cases, the target object is unmodified (i.e., the instructions are bare target).

In Fig. 5A, it is clear that people began fixating the target object (the apple) early in the utterance, so that by the time the word *apple* is uttered, the proportion of trials with a fixation on the apple is about 80%. This suggests that the object was of some interest even before it was mentioned, and also that there was no delay between hearing the object name and identifying the appropriate real-world referent (Allopena, Magnuson, & Tanenhaus, 1998; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). But notice that when the location to which the target object had to be moved matched its current location, participants seemed reluctant to stop looking at the target object, as indicated by significantly more looks to the target object beginning at the 900 ms time window and continuing through the 2100 ms window (see Table 5). Thus, participants spent more time looking at the target object in the matching condition than they did in the different condition.

Let us now turn to fixations on the correct location (Fig. 5B) in these same two conditions. Immediately upon hearing the word corresponding to the correct location (*box* or *towel*), participants looked to the corresponding location in the visual world. But at the offset of the utterance, looks to the correct location in the different condition were significantly more frequent than in the matching condition, beginning during the 600 ms time window and continuing until the 2100 ms window.

In summary, then, when people were asked to move the target object from one location to another of the same type, they perseverated in their looks to the target object. In addition, they took longer to begin looking at the appropriate new location, and the likelihood of fixating the location was not as high even a full two seconds after the utterance had finished. Clearly, it took longer for participants to commit to an interpretation or to figure out what the correct move was if they were asked to *put the apple on the towel* when that apple was already on a towel. This result is consistent with the findings from Experiments 1 and 2, which demonstrated that a speaker would never use such an instruction and that people judge such an instruction to be infelicitous relative to its context.

Effect of over-describing the target object: "*Put the apple in the box*" *versus* "*Put the apple on the towel in the box*". Fig. 6A shows looks to the target object for both the concise description (*put the apple in the box*) and for the over-described instruction (*put the apple on the towel in the box*) in the different conditions. Fig. 6B shows looks to the correct goal (e.g., the box) for the same instructions. Let us begin with looks to the target object (e.g., the apple)—Fig. 6A. In both conditions, the proportion of trials on which the apple is fixated increases immediately when the word *apple* is heard. Fixations continue to be high until about 1200 ms after utterance onset in the concise condition and 1800 ms in the over-described condition (see Table 5). This perseveration of looks to the target in the over-described condition is consistent with what we described in the previous analysis: When people hear *the apple on the towel*, they spend more time looking at the apple than they do when there is no mention of the towel.

Table 5
Results comparing fixation proportions ($n = 20$), by subjects, items, and minF' Experiment 3

Time bin	By subjects	By items	MinF'
Looks to target object in the two bare target instructions			
900	$F(1,19) = 5.15^{*}$	$F2(1,5) = 14.44^{*}$	$F3(1,24) = 3.80$
1200	$F1(1,19) = 76.04^{**}$	$F2(1,5) = 1180.61^{**}$	$F3(1,21) = 71.43^{**}$
1500	$F1(1,19) = 70.06^{**}$	$F2(1,5) = 74.82^{**}$	$F3(1,16) = 36.18^{**}$
1800	$F1(1,16) = 40.58^{**}$	$F2(1,5) = 8.64^{*}$	$F3(1,7) = 7.13^{*}$
2100	$F1(1,11) = 42.9^{**}$	$F2(1,5) = 12.39^{*}$	$F3(1,8) = 9.61^{*}$
Looks to correct location in the two bare target instructions			
600	$F1(1,19) = 12.67^{**}$	$F2(1,5) = 18.84^{**}$	$F3(1,20) = 7.58^{*}$
900	$F1(1,19) = 77.44^{**}$	$F2(1,5) = 72.25^{**}$	$F3(1,15) = 37.37^{**}$
1200	$F1(1,19) = 95.06^{**}$	$F2(1,5) = 67.72^{**}$	$F3(1,13) = 40.22^{**}$
1500	$F1(1,19) = 46.38^{**}$	$F2(1,5) = 12.39^{*}$	$F3(1,8) = 9.78^{*}$
1800	$F1(1,16) = 29.92^{**}$	$F2(1,5) = 7.34^{*}$	$F3(1,8) = 5.9^{*}$
2100	$F1(1,11) = 18.75^{**}$	$F2(1,5) = 4.41$	$F3(1,7) = 3.57$
Looks to target object between the two different location instructions			
1200	$F1(1,19) = 29.81^{**}$	$F2(1,5) = 159.26^{**}$	$F3(1,24) = 25.11^{**}$
1500	$F1(1,19) = 15.29^{**}$	$F2(1,5) = 36.6^{**}$	$F3(1,23) = 10.78^{**}$
1800	$F1(1,16) = 7.51^{*}$	$F2(1,5) = 12.25^{*}$	$F3(1,19) = 4.65^{*}$
Looks to the correct location between the two different instructions			
Box/towel	$F1(1,19) = 17.06^{**}$	$F2(1,5) = 13.32^{*}$	$F3(1,14) = 7.48^{*}$
900	$F1(1,19) = 19.98^{**}$	$F2(1,5) = 9.42^{*}$	$F3(1,10) = 6.4^{*}$
1200	$F1(1,19) = 36.6^{**}$	$F2(1,5) = 28.3^{**}$	$F3(1,14) = 15.96^{**}$
1500	$F1(1,19) = 12.11^{*}$	$F2(1,5) = 16.97^{**}$	$F3(1,19) = 7.07^{*}$
1800	$F1(1,16) = 5.62^{*}$	$F2(1,5) = 4.54$	$F3(1,14) = 2.51$
Looks to the incorrect location between the two different instructions			
Box/towel	$F1(1,19) = 11.16^{**}$	$F2(1,5) = 19.66^{*}$	$F3(1,21) = 7.12^{*}$
300	$F1(1,19) = 11.16^{**}$	$F2(1,5) = 9.18^{*}$	$F3(1,14) = 5.04^{*}$
600	$F1(1,19) = 8.82^{*}$	$F2(1,5) = 9.18^{*}$	$F3(1,16) = 4.5^{*}$
900	$F1(1,19) = 5.66^{*}$	$F2(1,5) = 8.76^{*}$	$F3(1,20) = 3.44$
1200	$F1(1,19) = 7.13^{*}$	$F2(1,5) = 5.38$	$F3(1,13) = 3.07$
1500	$F1(1,19) = 5.52^{*}$	$F2(1,5) = 3.28$	$F3(1,12) = 2.06$
Looks to target object between the two matching location instructions			
900	$F1(1,19) = 4.75^{*}$	$F2(1,5) = 4.49$	$F3(1,15) = 2.31$
1200	$F1(1,18) = 9.12^{**}$	$F2(1,5) = 8.47^{*}$	$F3(1,15) = 4.39$
1500	$F1(1,18) = 10.82^{**}$	$F2(1,5) = 4.24$	$F3(1,9) = 3.05$
Looks to the correct location between the two matching instructions			
600	$F1(1,19) = 4.54^{*}$	$F2(1,5) = 3.31$	$F3(1,13) = 1.91$
900	$F1(1,19) = 17.98^{**}$	$F2(1,5) = 9.0^{*}$	$F3(1,11) = 6.0^{*}$
1200	$F1(1,18) = 21.62^{**}$	$F2(1,5) = 5.02$	$F3(1,17) = 4.07$
1500	$F1(1,18) = 16.16^{**}$	$F2(1,5) = 3.39$	$F3(1,7) = 2.80$
1800	$F1(1,18) = 8.47^{**}$	$F2(1,5) = 2.07$	$F3(1,8) = 1.66$

* $p < .05$.
** $p < .01$.

Next, let us consider looks to the correct location (the box)—Fig. 6B. In the concise condition, looks to the box rise dramatically at *box*, drop a bit, and then stay high throughout all time windows. The pattern is similar in the over-described condition, except that looks to the box are delayed by about 300 ms. In addition, looks rise more slowly, so that even at 1800 ms post-utterance offset, the proportion of trials with looks to the box is lower in the over-described condition than in the concise condition.

Finally, consider looks to the empty towel (see Fig. 6C). In the bare target condition, there are essentially none, which is expected because no towel is mentioned in the utterance. In the over-described condition, fixations to the towel rise right after *on*, rise further at the word *towel*, and then drop when the second prepositional phrase begins. Thus, we see here precisely the pattern that Spivey et al. (2002) reported: when listeners hear *on the towel*, they look at the empty towel, and they only stop when

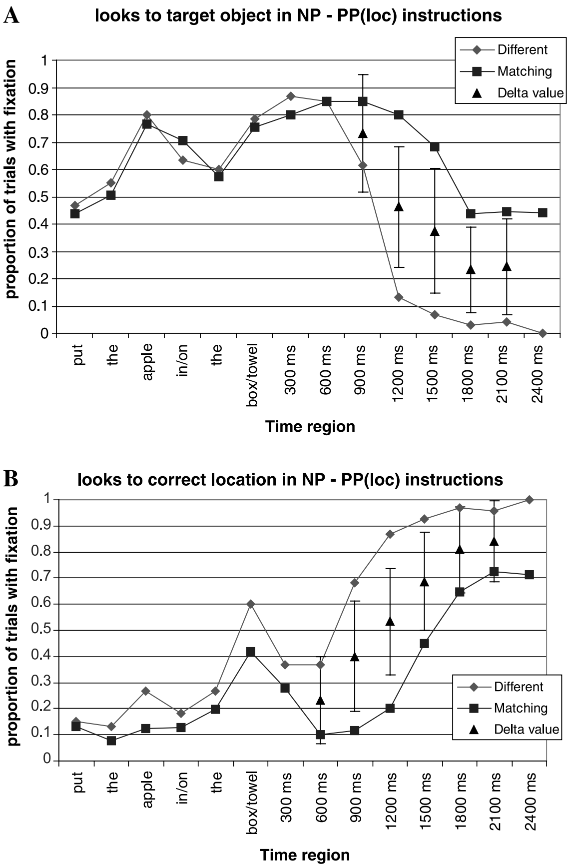

Fig. 5. (A) Momentary probability graphs comparing looks to the target object in the two bare target instructions in Experiment 3. (B) Momentary probability graphs comparing looks to the correct location in the two bare target instructions in Experiment 3.

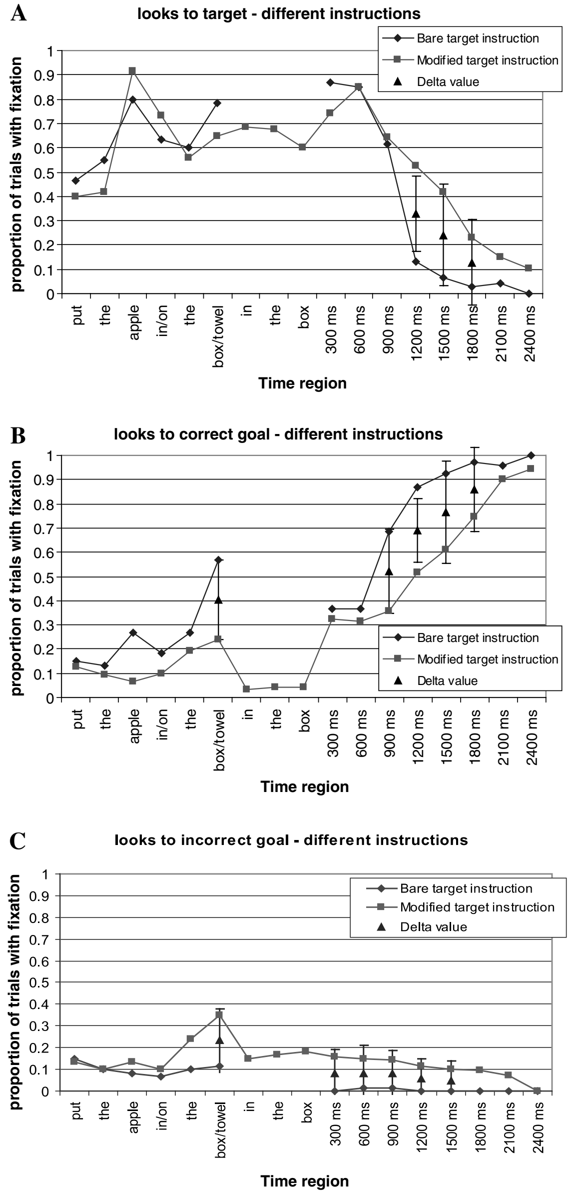

Fig. 6. (A) Momentary probability graphs comparing looks to the target object in the two different location instructions in Experiment 3. (B) Momentary probability graphs comparing looks to the correct location in the two different location instructions in Experiment 3. (C) Momentary probability graphs comparing looks to the incorrect location in the two different location instructions in Experiment 3.

they encounter the second prepositional phrase which indicates that *on the towel* is a modifier and not a location. This confusion leads to a delay in fixating the correct location, the box, in the over-described condition.

The cost of over-describing, then, is evident in an implicit measure of performance, namely eye-movement behavior. Even though participants explicitly indicated that they do not prefer concise descriptions over ones that are over-described (Experiment 2), their actual performance on a relevant task reveals that over-descriptions cause momentary confusion. Interestingly, listeners seem to be entirely unaware of this cost, perhaps because eye movements are generally made unconsciously, and recovery from the confusion is very fast. Performance on the task was near perfect: people virtually always moved the apple to the box. In addition, we observed that under-describing a location is costly in the condition in which participants heard *put the apple on the towel in the box*, as we reported in the comparison of the two concise instructions: when people hear *the apple on the towel*, they continue to look at the apple/towel longer than they do when there is no mention of the towel. This tendency again suggests that even though the comprehension system might momentarily consider the towel as a potential location for the apple, it is also

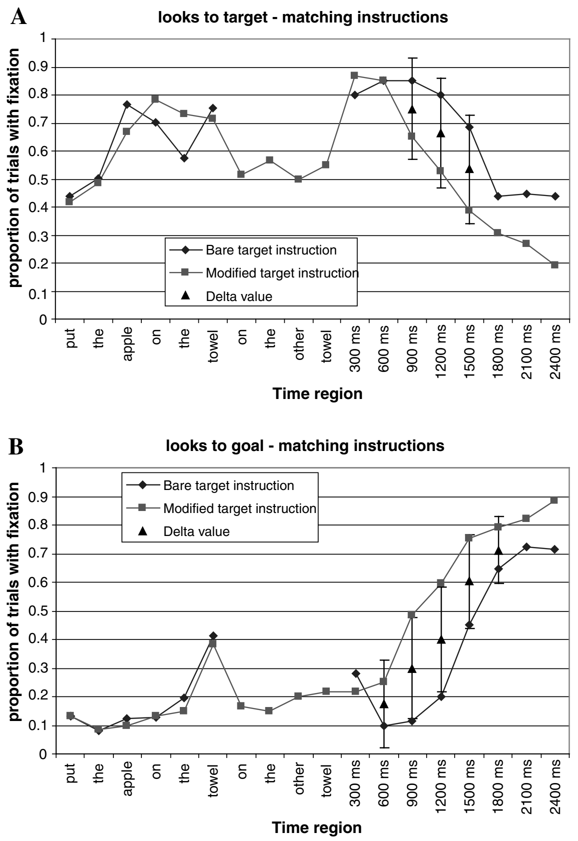

Fig. 7. (A) Momentary probability graphs comparing looks to the target object in the two matching location instructions in Experiment 3. (B) Momentary probability graphs comparing looks to the correct location in the two matching location instructions in Experiment 3.

somewhat confused about moving the object that is already on a towel.

Effect of under-describing the location and over-describing the target object: "*Put the apple on the towel*" *versus* "*Put the apple on the towel on the other towel*". Fig. 7A shows looks to the target object for both the concise (*put the apple on the towel*) and for the over-described instruction (*put the apple on the towel on the other towel*) in the matching conditions. Fig. 7B shows looks to the correct goal (e.g., the towel) for these same instructions.

Let us begin with looks to the target object, the apple (see Fig. 7A). Again, the proportion of trials on which the apple is fixated spikes immediately when the word *apple* is heard. But eye-movement behavior differs in that even at 1500 ms past utterance offset, people are more likely to be looking at the apple/towel when they received the simple instruction *put the apple on the towel* compared to the instruction with the unnecessary prepositional phrase modifier. Again, this is the cost for under-specifying the location (see Table 5).

Now let us consider looks to the empty towel (see Fig. 7B). In both conditions, those looks increase quickly when the word *towel* is heard. Interestingly, they dramatically fall again before rising during the time intervals after utterance offset. The second rise in looks occurs later in the bare target condition—at about 1200 ms post-offset, compared to 600 ms post-offset in the modified target condition. This pattern supports the idea that when people first hear the word *towel*, they look at the corresponding object. It is only after the utterance is over that they realize that towel is the location, and they refixate the towel in order to move the apple to it.

General discussion

The experiments described above focused on the Gricean Maxim of Quantity and the extent to which it captures normal linguistic behavior. The Maxim of Quantity has two components: the first is that speakers should make their contributions as informative as is required for successful communication; listeners expect linguistic expressions to allow them to identify a unique referent. The second component is that speakers should make contributions no more informative than necessary; listeners will assume that information provided is relevant, and so they will either interpret a complex noun phrase to indicate that a set of objects exists, or they will find the complex expression infelicitous, because it over-describes.

These experiments provide only limited support for the psychological reality of this Gricean principle. We observed that speakers tend to provide enough information to allow a unique referent to be identified, as the first experiment showed that complex noun phrases were produced when there was more than one of the relevant object type in the visual world. In addition, listeners judging the quality of instructions consider an unmodified noun phrase in the context of more than one relevant object to be infelicitous. Thus, speakers and listeners avoid under-descriptions that might lead to confusion about referent identity (Ferreira et al., 2005).

A similar type of under-description that speakers avoid and listeners dislike is one that fails to distinguish two identical locations. Specifically, we found that speakers never simply say *put the apple on the towel* when the apple is already on a towel and it is to be moved to another one. Moreover, listeners judge this sort of instruction to be the most infelicitous of all the instruction types we examined. This pattern is interesting, because the problem with calling the towel simply *the towel* rather than *the other towel* is not that it creates confusion or ambiguity about which towel is the right one, because only if the apple is placed on another towel can it be correctly described as having been moved. This situation creates less ambiguity than one in which an apple is simply called *the apple* when more than one exists in the relevant world, and yet this expression bothered speakers and judges less than the unmodified towel. In the third experiment we observed that this expression did confuse listeners: we saw that when they heard *put the apple on the towel* they continued to look at the apple throughout the instruction and even after it had been completed, as if they were somewhat puzzled by what they were being asked to do. We observed this pattern of continued fixations on the apple throughout the trial regardless of whether the word *towel* was followed by the real location or whether the utterance ended there.

Now let us consider over-descriptions, which, as we pointed out at the beginning of this report, have played an important role in helping psycholinguists understand the architecture of the language processing system. Is an expression such as *the apple on the towel* produced only when there is more than one apple, and do listeners dislike complex expressions in contexts that are hypothesized not to support their use? Our experiments reveal a complex picture. On one hand, the first experiment showed that over-descriptions are produced about one-third of the time by naïve speakers. In addition, listeners judging the quality of instructions failed to rate over-descriptions as worse than concise ones, thus providing no support for the idea that over-descriptions are bad. But, interestingly, we found that listeners' implicit behavior did indeed reveal the cost associated with over-describing an object. We found that when people heard *put the apple on the towel in the box*, they looked at the empty towel immediately upon hearing the corresponding lexical item, and they were somewhat delayed in their looks to the correct location (the box). When we put this finding together with the other eye-movement results showing that the expression *apple on the towel* leads to perseveration in looks to the apple, we see that this combination of instruction and visual world leads to less efficient performance, at least compared to the condition in which the expression is appropriately concise and the listener is not asked to make the pragmatically odd action of moving an apple from one towel to another (i.e., the best instruction is *put the apple in the box*).

The next question is, what does this inefficient performance reflect? The confusion caused by simply asking a participant to put the apple on the towel when the apple is already on a towel is fairly easy to explain: As one participant pointed out to us during the experiment, the apple is already on a towel, and so it is pragmatically necessary to acknowledge that state of affairs. The location towel must be referred to as *the other towel* or *the opposite towel* to distinguish it from the one it is on and to thereby provide a rationale for having it moved. But why do listeners look at the towel when they hear *put the apple on the towel* (whether the utterance ends there or continues) in the matching location conditions? We have seen that they are unlikely to have ever heard such an instruction in the context of an object that is already on a towel, and they judge such a combination of instruction and context to be infelicitous. Why, then, do they immediately make eye movements to the empty towel?

We believe there are two general classes of explanation. The first is that when the parser encounters *put the apple on the towel* (...), it builds a syntactic structure in which the prepositional phrase *on the towel* is a location rather than a modifier. Three different mechanisms can explain this tendency for the simple structure to be built initially. The first is the one offered by the Referential Model: the Garden-Path occurs because the one-referent context does not support complex modification, and the principle of parsimony favors interpretations that require fewer presuppositions (Altmann & Steedman, 1988; Crain & Steedman, 1985; Spivey et al., 2002; Tanenhaus et al., 1995). But this account of why the parser builds the simpler syntactic structure is difficult to reconcile with the results of the first and second experiments, because those showed that speakers will often produce an over-described utterance and that listeners do not in fact judge such an utterance to be infelicitous. Moreover, we have seen that a speaker would likely never produce the instruction *put the apple on the towel* in the relevant visual context, and listeners view it as pragmatically unacceptable. It would appear, then, that it is the simple instruction that is unsupported by the context. Thus, for the explanation in terms of the Referential Model to go through, there has to be an additional story about why the comprehension system would be sensitive to one type of pragmatic constraint, the putative bias against over-descriptions, but insensitive to another, the apparent prohibition against under-describing a location when it is of the same type as the target object's current location. Given that the latter constraint seems much stronger, as indicated by production data and judgments, it is difficult to see why it is more defeasible. It appears to us that the explanation which appeals to the Referential Model is implausible. It is also important to note that the constraint-based, lexicalist version of the Referential Model (Spivey et al., 2002; Tanenhaus et al., 1995) assumes that frequency of exposure affects how easily linguistic expressions can be processed. Our finding that the simple instruction *put the apple on the towel* has probably never been encountered in a context in which the apple is already on a towel and is to be moved to another towel leads to the expectation that the location interpretation of *on the towel* would be ruled out by a parser sensitive to this type of distributional information. But clearly, this is not what we observed, nor has this been found in previous investigations of this particular question (Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999).

The second mechanism that might cause the parser to build a simple syntactic structure initially is Minimal Attachment, as is assumed in the Garden-Path Model of parsing (Frazier, 1978). The parser initially makes the prepositional phrase a location because that structure requires the postulation of fewer nodes than a structure in which the prepositional phrase is a nominal modifier. The Garden-Path Model assumes that the parser is informationally encapsulated, and we therefore expect this mechanism to operate even though the instruction *put the apple on the towel* is pragmatically odd and probably never has been heard. The Minimal Attachment preference can be viewed as a fast and frugal heuristic (Gigerenzer & Goldstein, 1996) that the comprehension system applies to input to arrive at an initial analysis quickly and without the need to consult much information. Thus, the Garden-Path Model can account for the tendency to make looks to the empty towel when the instruction is heard, even though the simple instruction is odd and even though listeners probably have never heard it in the relevant context.

The third mechanism that can explain the hypothesized tendency of the comprehension system to build the location analysis of the ambiguous prepositional phrase is that verb argument structure is used to build its initial syntactic structure rather than Minimal Attachment. The verb *put* was used in all the experiments reported here and in previous work, but *put* has an important property: a location-prepositional phrase is obligatory. Thus, if the parser works so that it attempts to saturate its argument structures as quickly as possible (Pritchett, 1992), or if the parser's initial analysis corresponds to the most frequently used argument structure (Boland, Tanenhaus, Garnsey, & Carlson, 1995; MacDonald et al., 1994), then we would expect to find the prepositional phrase interpreted initially as a location, and therefore we would observe listeners making looks to the empty location in the visual world we created. Again, however, if the explanation appeals to argument structure frequency, then something more needs to be said to explain why this frequency information is influential but the non-occurrence of the instruction itself is not. Our conclusion, then, is that the most likely account of our finding that listeners looked at the empty towel even though the instruction *put the apple on the towel* is pragmatically infelicitous and would probably never have been encountered previously in the relevant visual context is that the comprehension system initially uses Minimal Attachment and verb argument structure information to build its initial analysis. Moreover, the system does not take into account any other potentially relevant information, including the pragmatic implausibility of the instruction-context combination.

Let us now return to the question of what these experiments tell us about the Gricean Maxim of Quantity. We have argued that the Maxim of Quantity is only partially supported by our results: under-descriptions are indeed bad, but over-descriptions are often produced. But given that an over-description contains more words than a properly concise description, why would over-descriptions be generated so often (not only in Experiment 1 but also in Deutsch & Pechmann, 1982)?

We argue that the way to approach this question is to consider the situation from the perspective of the speaker. The speaker is confronted with a visual context consisting of four objects: an empty towel, an empty box, a frog by itself, and an apple on a towel (for example). Notice that one object is clearly special—the apple on the towel. It is the only object that is actually a combination of two—an apple and a towel together. Moreover, the speaker can see that there is a second towel (an empty one) in the relevant visual world. In this situation, then, the apple on the towel is likely a salient entity. Recall that in Experiment 3, listeners looked at the apple-towel object more than the others even during the word *put*, indicating that the object was already of some interest even before its name was encountered. In the speaker's mental representation of the visual context, then, the apple is an apple on a towel, and they are likely (30% of the time) to refer to it as such. We suggest that it would take additional mental operations for the speaker to provide a concise description, because he or she would have to (1) recognize that the modifier is not necessary from the perspective of the listener's ability to identify the correct referent, and (2) delete the modifier from the representation that drives the ultimate form of the utterance. Previous work has shown that speakers are less-than-perfect at modeling the minds of their listeners (Arnold, Wasow, Asudeh, & Alrenga, 2004; Ferreira & Dell, 2000; Ferreira et al., 2005; Snedeker & Trueswell, 2003), and some studies suggest that mental effort is required to avoid egocentricity (Keysar, Barr, & Horton, 1998). Indeed, this analysis of speakers' needs leads us to give our speakers credit for being able to adjust their representations and utterances the majority of the time. But, by the same token, on this view we expect that speakers will produce over-descriptions: they will do so when the linguistic expression captures the relevant situation from their point of view, and they do not engage in the processes necessary to adjust it to make it ideal for the audience.

This analysis of speakers' over-descriptions treats the question with which we began this paper in a way that is now familiar from the literature on language production. The question we posed initially is what determines whether the speaker refers to an apple as *the apple*, *the ripe apple, the apple on the towel*, and so on? We suggest that this issue be viewed as a version of the one that has been asked about speakers' choice from among syntactic options: what determines whether a speaker says *the boy ate the apple* or *the apple was eaten by the boy*? Experimental work on language production has demonstrated that the availability of concepts in the speaker's mental representation of an event affects utterance form because available concepts will tend to occupy sentence-initial position, and the rest of the structure will accommodate that placement using relevant linguistic constraints on word order (Bock, 1987; Christianson & Ferreira, 2005). People, then, will say *the apple was eaten by the boy* when the concept APPLE is active (e.g., it is the topic), resulting in the creation of a passive. Similarly, people will tend to refer to the apple as *the apple on the towel* when that entity as a complex object is salient. In both cases, the speaker might adjust the utterance to accommodate listener needs, but this operation does not work invariably or infallibly.

This then leads us to our final question, which is why do listeners judging utterances appear to like over-descriptions, at least as much as they like concise descriptions? An explanation we can dismiss is that they are simply insensitive to pragmatic infelicity in the judgment paradigm we used in the second experiment. This idea is implausible because we saw that other types of utterances were judged as infelicitous, especially ones in which *the towel* was not modified (and this can be viewed as a fairly subtle pragmatic anomaly, given that it has been overlooked in the literature until now). We believe that listeners are confronted with over-descriptions often enough that they are not troubled when they encounter them. Indeed, there is other evidence that listeners do not always assume contrastive interpretations of modifiers (Sedivy, 2003), which leads us to expect that over-descriptions would not necessarily invoke a set of objects and thus infelicity when the set does not exist. Moreover, in the judgment task, participants could very well employ a heuristic of judging more explicit utterances to be better than concise ones, simply because they contain more content, and this heuristic would favor over-descriptions. It is true that if those same judges had to execute the instructions rather than simply evaluating them, then they likely would experience some confusion (viz., they would make eye movements to the empty towel), but this cost is not something of which they are aware. The lack of awareness is probably related to the fact that the "price" they would pay is that they would make unnecessary eye movements, but these are typically made unconsciously, and so the comprehender probably does not even realize he/she has looked at the empty towel when it was not relevant to executing the instruction.

We conclude that the language processing system is only moderately Gricean, at least with respect to the Maxim of Quantity. Speakers often violate this Maxim by producing over-descriptions (and often under-descriptions as well; Ferreira et al., 2005), and over-descriptions do not always bother listeners. Given our findings and this conclusion, we believe that it is necessary to re-consider the widespread tendency to invoke this Gricean Maxim to explain phenomena of language processing. This tendency can manifest itself in decisions about experimental designs; for example, Snedeker and Trueswell (2003) did not include conditions which would have constituted over-descriptions for fear that their participants would view them as pragmatically odd. But it can also manifest itself as an assumption that motivates an entire architecture for processing, as in the Construal Model of comprehension (Frazier & Rayner, 1982; see also Brysbaert & Mitchell, 1996) or, more directly, in the Referential Model (Altmann & Steedman, 1988; Crain & Steedman, 1985). Indeed, it would be almost impossible to catalogue all the psycholinguistic explanations which have appealed to some version of the idea that listeners have difficulty with construction X because listeners believe the speaker would have produced Y instead given the Gricean Maxim of Quantity. Explanations of this sort are not necessarily wrong, but we believe our data give us cause to scrutinize them carefully, because speakers and listeners are only moderately Gricean in their adherence to the Maxim of Quantity.

Acknowledgments

The authors would like to thank John Henderson and John Hale for their guidance on this project. Thanks also to Emily France, Ritik Tiwari, and Brett Guth for their help running the experiments and preparing the stimuli. Paul E. Engelhardt was supported by NSF-IGERT Grant DGE-0114378 and NIMH Grant MH-65310.

References

Allopena, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439.

Altmann, G. T., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action* (pp. 347–386). New York: Psychology Press.

Altmann, G. T., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191–238.

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language, 51*, 55–70.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). Englewood Cliffs, NJ: Prentice-Hall.

Bock, J. K. (1987). Coordinating words and syntax in speech plans. In A. Ellis (Ed.), *Progress in the psychology of language* (pp. 337–390). Hillsdale, NJ: Erlbaum.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language, 34*, 774–806.

Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *The Quarterly Journal of Experimental Psychology, 49A*, 664–695.

Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition, 98*, 105–135.

Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.

Clifton, C., & Ferreira, F. (1989). Ambiguity in context. *Language and Cognitive Processes, 4*, 77–103.

Cooper, R. M. (1974). The control of eye fixation by the meaning of the spoken language. *Cognitive Psychology, 6*, 84–107.

Crain, S., & Steedman, M. J. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives*. Cambridge: Cambridge University Press.

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition, 11*, 159–184.

Eberhard, K., Spivey-Knowlton, M., Sedivy, J., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*, 409–436.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368.

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

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition, 96*, 263–284.

Frazier, L. (1978). On comprehending sentences: Syntactic parsing strategies. Ph. D. Thesis, University of Connecticut. Indiana Linguistic Club.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650–669.

Grice, P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. III, pp. 41–58).

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Sciences, 7*, 46–50.

Levison, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicatures*. Cambridge, MA: MIT Press.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.

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

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

Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago, IL: The University of Chicago Press.

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., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual interpretation. *Cognition, 71*, 109–147.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1521–1543.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45*, 447–481.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*, 103–130.

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.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of cognition and perception*. San Diego, CA: Academic Press.

Trueswell, J. C., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence comprehension in young children. *Cognition, 73*, 89–134.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285–318.