Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
COGNITION
Cognition 96 (2005) 263–284
www.elsevier.com/locate/COGNIT

How do speakers avoid ambiguous linguistic expressions?

Victor S. Ferreira[a,*], L. Robert Slevc[a], Erin S. Rogers[b]

[a]*Department of Psychology 0109, University of California, San Diego, La Jolla, CA 92093-0109, USA*
[b]*University of California, Los Angeles, CA, USA*

Received 19 January 2004; revised 10 May 2004; accepted 9 September 2004

Abstract

Three experiments assessed how speakers avoid linguistically and nonlinguistically ambiguous expressions. Speakers described target objects (a flying mammal, *bat*) in contexts including foil objects that caused linguistic (a baseball bat) and nonlinguistic (a larger flying mammal) ambiguity. Speakers sometimes avoided linguistic-ambiguity, and they did so equally regardless of whether they also were about to describe foils. This suggests that comprehension processes can sometimes detect linguistic-ambiguity before producing it. However, once produced, speakers consistently avoided using the same linguistically ambiguous expression again for a different meaning. This suggests that production processes can successfully detect linguistic-ambiguity after-the-fact. Speakers almost always avoided nonlinguistic-ambiguity. Thus, production processes are especially sensitive to nonlinguistic- but not linguistic-ambiguity, with the latter avoided consistently only once it is already articulated.

Keywords: Language production; Referential communication; Ambiguity; Monitoring; Homophones

For communication to succeed, speakers must be *cooperative*, in that they must cater features of their utterances to the needs of their addressees. For example, in a noisy environment, speakers must speak more loudly, or with an inexperienced interlocutor, speakers must use more common words and simpler phrases. But what cognitive mechanisms underlie speakers' ability to accommodate their listeners in this manner?

* Corresponding author. Tel.: +1 858 534 6303; fax: +1 858 534 7190.
E-mail address: vferreira@ucsd.edu (V.S. Ferreira).

0022-2860/$ - see front matter
doi:10.1016/j.cognition.2004.09.002

In this paper, we address the question of speaker cooperativity in the domain of *ambiguity avoidance*. Ambiguity is a fundamental threat to the comprehension process: As comprehenders aim to decode the meanings of speakers' utterances, they encounter difficulties when an utterance means more than one thing. Despite the threat of ambiguity, addressees understand speakers most of the time. Is this because speakers take steps to avoid ambiguity? Or is ambiguity dealt with some other way?

For example, take the word *bat*, which is a *homophone*—it can refer either to a flying mammal or to an instrument for hitting baseballs (or cricket balls). In contexts in which either interpretation is possible, the homophone description *bat* is therefore ambiguous and so should be avoided. This is an example of a *linguistic*-ambiguity, in which two independent meanings correspond to the same surface form because of some accident or limitation of linguistic encoding. Besides homophony, linguistic-ambiguity also arises because of segmentation ambiguity (*a back* vs. *aback*) and syntactic ambiguity (*stolen painting found by trees*).

Ambiguities can also be *nonlinguistic*. Nonlinguistic-ambiguities arise in contexts that include multiple instances of similar meanings, so that a single term that encompasses all of those instances is ambiguous. For example, in the context of two flying mammals, one larger and one smaller, the term *bat* is ambiguous and so should be avoided.

Consider how these kinds of ambiguity arise within standard models of language production. To produce a linguistic expression, speakers proceed through the sequence of processing stages shown in Fig. 1 (Bock, 1982; Dell, 1986; Garrett, 1975; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999; for a somewhat different view, see Caramazza, 1997).

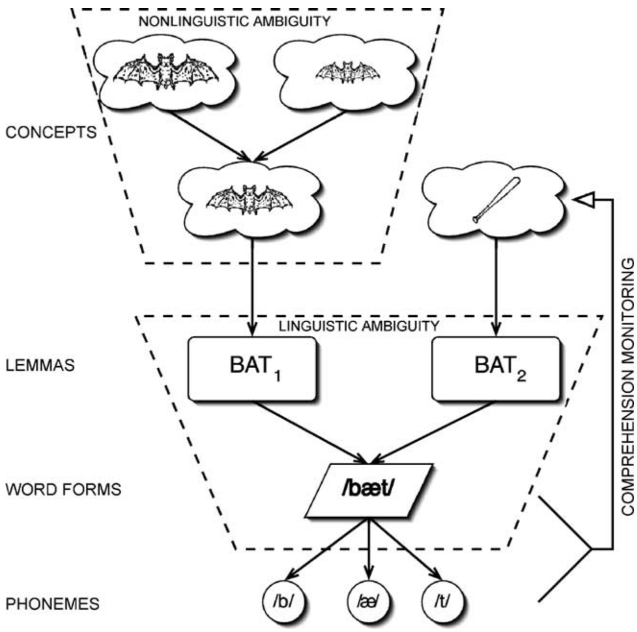

Fig. 1. The representation of nonlinguistic- and linguistic-ambiguity within the processing stages of word production.

Production begins with a meaning or conceptual representation of the idea a speaker wants to express. The production system uses this meaning to retrieve syntactically based lexical representations termed *lemmas*, which in turn are used to retrieve phonologically based lexical representations termed *word forms* (or sometimes *lexemes*). Word forms are then used to retrieve segmental and phonological information en route to articulation.

Within this model, linguistic and nonlinguistic-ambiguities arise at different processing stages (see Fig. 1). Nonlinguistic-ambiguities arise because related specific meanings map to a common more general meaning, and that common more general meaning maps to a single linguistic representation (one lemma). In contrast, linguistic ambiguities arise because distinct meanings map to distinct syntactic (lemma) representations and then to a single phonological (word) form. Therefore, the similarity (identity) that underlies nonlinguistic-ambiguity is represented at the level of meaning, whereas the similarity (identity) that underlies linguistic-ambiguity is represented at the level of linguistic (phonological) form.

Note that a consequence of the different representational bases of nonlinguistic- vs. linguistic-ambiguity is that each is likely to be avoided by different processing strategies. On the one hand, nonlinguistic-ambiguities can be avoided in the normal course of formulating linguistic expressions. That is, because the similarity that underlies nonlinguistic-ambiguity is represented at the level of meaning, that similarity is available to influence the very first steps of production, just like any other feature of meaning. This predicts that when speakers are confronted with nonlinguistic-ambiguity, they ought to avoid it consistently, and in fact they do. Research using referential communication tasks has revealed that when speakers are instructed to describe objects that contrast with other nearby objects in size or color, they distinguish the two (e.g. by saying *large bat* rather than *bat*), sometimes doing so even when they know their addressees do not know about the contrasting objects (thereby hindering addressees' performance; e.g. Horton & Keysar, 1996; Nadig & Sedivy, 2002; Wardlow & Ferreira, 2003). Similarly, when speakers describe two referents from the same basic-level category (e.g. shoe), they consistently use less preferred subordinate labels (e.g. sneaker), and they continue to do so with the same listener even when the labels are no longer distinctive (Brennan & Clark, 1996).

On the other hand, avoiding linguistic ambiguities is trickier, because the similarity that underlies linguistic-ambiguity does not become available until after the production process itself has begun. Given this, speakers might use two general strategies to detect (and then avoid) linguistic-ambiguity. One is via a process also illustrated in Fig. 1, which is termed *comprehension monitoring* (or sometimes, *perceptual-loop monitoring*). That is, many models of production (e.g. Hartsuiker & Kolk, 2001; Levelt, 1989; Postma, 2000) include a final stage of processing during which speakers comprehend and evaluate their to-be-produced (inner) speech for adequacy on multiple dimensions. For example, to detect speech errors—unintended material that results from the derailment of language-production processes—speakers might literally comprehend their internal speech and recognize that it does not match the meaning they wish to express. A similar strategy could operate to detect linguistic-ambiguity: As speakers produce utterances, they may comprehend those utterances for whether they can mean more than one thing in their contexts, and if so, take steps to avoid the detected ambiguity (Levelt, 1989, raises this possibility).

A different way that linguistic-ambiguity might be avoided is primarily by the action of production processes rather than comprehension processes. Specifically, to detect ambiguity, speakers might apply production processes (at least to some extent) to multiple meanings in the current context. If in doing so, production processes retrieve the same word form for two distinct meanings, then by themselves those production processes will have revealed ambiguity.

To illustrate these strategies for linguistic-ambiguity avoidance, imagine that speakers are presented with a display that includes pictures of a flying mammal and a baseball bat, and are instructed to uniquely describe the flying mammal. According to a comprehension-based account, as speakers retrieve the word form */bæt/*, they comprehend the intended as well as the unintended meaning of that word form, recognize that the unintended meaning is represented in the current context, and so take steps to avoid the ambiguity. According to a production-based account, ambiguity can be detected if speakers retrieve a word-form for the flying mammal (*/bæt/*), a word-form for the baseball bat (*/bæt/*), and then recognize that the two are the same word-form. Consequently, the production-based account claims that speakers must at least to some extent apply production processes to each meaning to detect the ambiguity. Note that the critical difference between the comprehension- and production-based accounts is how an ambiguous form (*/bæt/*) is connected to an unintended meaning (the baseball bat): According to the comprehension-based account, this is done via comprehension processes (from the ambiguous form, up via comprehension monitoring, to the unintended meaning), whereas according to the production-based account, this is done via production processes (from the unintended meaning, down through production processes, to the ambiguous form). (Note also that the production-based account may still use monitoring processes to detect that the two retrieved word forms are in fact identical. Hence, the critical difference between the accounts is not simply the involvement of monitoring, but rather, whether it is monitoring processes or production processes that draw the connection between ambiguous form and unintended meaning.).

There is a third possibility: Speakers may not be particularly sensitive to the linguistic-ambiguity of their expressions at all. Indeed, past research on linguistic-ambiguity avoidance suggests this possibility. Nearly all of this research has investigated the avoidance of syntactic ambiguity specifically, to determine whether speakers phrase sentences or use prosodic features to disambiguate otherwise ambiguous structures. Most investigations have shown either that speakers do not distinguish ambiguous structures (Allbritton, McKoon, & Ratcliff, 1996; Craig, Nicol, & Barss, 1995; Elsness, 1984; Ferreira & Dell, 2000; Mims & Trueswell, 1999), or that if they do, they do so even in unambiguous contexts (suggesting that speakers did not avoid ambiguity per se; Kraljic & Brennan, 2003; Schafer, Speer, Warren, & White, 2000). Only one study has shown that speakers use prosody to avoid syntactically ambiguous structures only in ambiguous situations (Snedeker & Trueswell, 2003). Of course, the failure to consistently observe syntactic-ambiguity avoidance may not speak to the avoidance of linguistic ambiguities in general, but instead may only reflect challenges specific to the avoidance of syntactic ambiguity, such as the complex and abstract nature of syntactic structures.

Together, these observations suggest first that speakers should effectively avoid nonlinguistic-ambiguity, as it can be avoided in the normal course of language production.

However, speakers may or may not effectively avoid linguistic-ambiguity, as special comprehension- or production-based strategies are likely necessary for detecting it. Here, we present three experiments that used a referential-communication task (which has revealed successful nonlinguistic-ambiguity avoidance) to assess both nonlinguistic- and linguistic-ambiguity avoidance. We expected that speakers would successfully avoid nonlinguistic-ambiguity; of new interest is whether they would also avoid linguistic ambiguities under similar circumstances, and if so, whether through a comprehension- or a production-based strategy.

In the experiments, speakers were shown displays like those in Fig. 2, which included four objects: a target object (labeled with a dot numbered "2" in Fig. 2), which on critical trials could be described with a homophone label; a foil object (labeled with a dot numbered "3" in Fig. 2), which (sometimes) created some form of ambiguity; and two filler objects, which made the displays more complex (thereby discouraging speakers from adopting task-specific strategies). We asked speakers to uniquely describe target objects (e.g. the flying mammal) and measured how often they used potentially ambiguous bare homophone labels (e.g. *bat*). To assess nonlinguistic-ambiguity avoidance, speakers described displays like in Fig. 2a, in which the foil object (a larger flying mammal) came from the same conceptual category as the target object (a smaller flying mammal), but was describably different (e.g. in terms of size). To assess linguistic-ambiguity avoidance, speakers described displays like in Fig. 2b, in which the foil object (a baseball bat) had a name that was a homophone of the name of the target object (a flying mammal). Finally, as a control, speakers described the same target objects when the foil did not create any form of ambiguity at all, as in Fig. 2c (where the foil object is a goat). Because the control condition is unambiguous, it allows us to operationalize ambiguity avoidance as the extent to which speakers use bare homophone labels less in the ambiguous conditions (Fig. 2a and b), compared to in the unambiguous control condition (Fig. 2c).

To assess whether speakers avoid ambiguity with a comprehension-based or a production-based ambiguity-detection strategy, speakers were shown displays in which a dot appeared and moved from one picture to another in the manner shown by the numbers in Fig. 2 (the numbers themselves were not presented). Each display was tested in two conditions, which were identical except the instructions given to speakers.

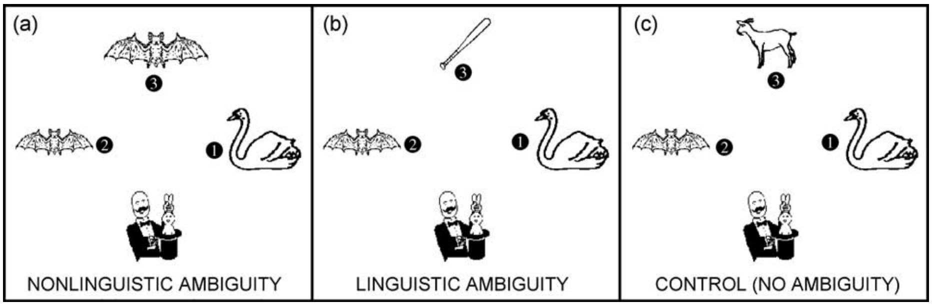

Fig. 2. Sample displays from Experiment 1. (a) Nonlinguistic-ambiguity condition, (b) linguistic-ambiguity condition, (c) control condition. In all cases, the target is indicated by the dot numbered "2" and the critical foil by the dot numbered "3".

With the *second-last* instruction, speakers were to name just the second-last picture that was indicated, which (on critical trials) was always the target object. With the *all* instruction, speakers were to name all indicated pictures, which (on critical trials) was a filler object, and then the target object, and then the foil object. Thus, with the *second-last* instruction, speakers applied production processes only to the target object, whereas with the *all* instruction, speakers applied production processes both to the target object and to the foil object (and in that order).

If speakers effectively avoid nonlinguistic- and linguistic-ambiguities, we should find that they produce fewer bare homophones in the nonlinguistic- and linguistic-ambiguity conditions, respectively, than in the unambiguous control condition. Furthermore, to the extent that speakers produce fewer bare homophones with the *all* than with the *second-last* instruction in the linguistic-ambiguity condition, it suggests that applying production processes to the ambiguity-causing foil improves ambiguity detection, which in turn implies that ambiguity detection happens via production processing. In contrast, if speakers do not produce fewer bare homophones with the *all* instruction in the linguistic-ambiguity condition, it suggests that applying production processes to the ambiguity-causing foil as well as the target provides no ambiguity-detection benefit, which implies that any linguistic-ambiguity avoidance that was observed was not due to a production-based strategy.

Experiment 1 manipulated one additional factor. Research on communication has shown that communicative behavior (e.g. avoiding ambiguity) can be sensitive to the presence of genuine addressees for whom communication is relevant (e.g. Lockridge & Brennan, 2002; Schober, 1993). This may be because speakers find it easier to model the needs of real addressees rather than hypothetical addressees, or because real addressees provide feedback that cues speakers to the existence of or an appropriate remedy for the ambiguity (Lockridge & Brennan, 2002). To explore the effect of the presence of a hypothetical vs. actual listener, Experiment 1 tested two groups of speakers. One group described displays to addressees who were given printed versions of the pictures arranged in a row. The addressees' task was to number their pictures in the same order that speakers described them. Speakers were asked to describe pictures so that addressees could accomplish this task. Addressees were genuine subjects, and the roles of speaker and addressee were randomly assigned at the beginning of the experiment with a coin flip. Speakers in the other group described displays to a hypothetical listener. If ambiguity avoidance is sensitive to the presence of actual addressees, qualitatively different ambiguity-avoidance should be observed between these two speaker groups.

1. Experiment 1

1.1. Method

Speakers. Seventy-two members of the University of California, San Diego community participated for class credit or cash payment ($6–$8). Of these, 48 were speakers and 24 were addressees paired with speakers. All speakers reported that English was their native language.

Apparatus. The experiment was run on Macintosh 6500/250 computers, with 17-in. CRT displays set to a resolution of 832×624 and a color depth of 256 colors. Voice responses were measured with PsyScope button boxes (Cohen, MacWhinney, Flatt, & Provost, 1993) and recorded with Marantz PMD201 cassette recorders. Speakers wore Shure unidirectional headset microphones, which fed input to both the response box and the tape recorder. Addressees in the real-addressee condition (see below) were given a booklet that showed the pictures in the experiment, four in a row on each page, representing each quadruplet that speakers were presented with. Below each picture was a box large enough to accommodate a written digit.

Materials and design. Materials were developed around the linguistic-ambiguity condition. Twenty-seven pairs of pictures were developed where each picture in the pair could be named with homophonic terms (e.g. a flying mammal and a baseball bat). Pictures were selected from Bates et al. (2003), Ferreira and Cutting (1997), Snodgrass and Vanderwart (1980), from internet searches, or were drawn for these experiments specifically.

The 27 pairs of pictures were normed on a separate group of 40 subjects. The pairs of pictures were compiled into two lists such that each list had one member of each pair. Lists were printed as booklets with nine pictures to a page. Each subject was given one of the lists (20 subjects per list), and was asked to write "the first name that came to mind" for each picture. We counted how often speakers provided the appropriate homophonic (i.e. ambiguous) target label for each of the 54 pictures. From these counts, we chose 18 critical picture-pairs to use in Experiments 1–3. For each critical picture, speakers used the targeted homophonic label at least 45% of the time. Sixteen of the critical pictures were described with the homophonic label at least 90% of the time. Overall, homophonic labels were provided an average of 79.3% of the time. The names of the critical pictures are shown in the Appendix.

One member of each of the critical pairs was designated as the target picture, and the other as the linguistically ambiguous foil. To maximize analyzable data, we chose the picture from each pair that was most often described with a homophonic label as the target picture.

For each target picture, we created a nonlinguistically ambiguous foil by modifying the line drawing of the target picture, or finding or generating another line drawing that was from the same conceptual category as the target object, but was describably different. For 11 of 18 targets, the nonlinguistically ambiguous foil differed in size; for 3, it differed in shading; for 2 it differed in numerosity; and for 2, it differed otherwise (square and round glasses, upper case and lower case letters). Then, for each target, a control foil was selected by randomly selecting a line drawing (from the above sources) that was neither conceptually nor descriptively similar to the target object. Finally, two more filler pictures were chosen for each target by randomly selecting two more line drawings with the same criteria as the control foils.

An additional 36 quadruplets of pictures were selected for all-filler trials. Half of the targets for these allowed one-word descriptions like *octopus* and *shoe* (to encourage simple descriptions like *bat*) and the other half allowed more complex descriptions like *movie projector* and *staple remover* (to encourage more complex description like *flying bat*). For each of these sets, two quadruplets included a nonlinguistically similar target object, and four more quadruplets had conceptually similar pairs of objects where neither was the target.

Overall, each speaker saw 54 quadruplets of pictures, of which 18 were critical. Of the 18 criticals, six were presented in the linguistically ambiguous condition (i.e. included the linguistically ambiguous foil), six in the nonlinguistically ambiguous condition, and six in the control condition. Of each of these six, three quadruplets were presented with the *second-last* instruction and three with the *all* instruction (see below). Both ambiguity and instruction were manipulated within speakers and within items (where each item is defined by a target picture) in counterbalanced fashion. In addition, 24 speakers were tested with real addressees (see below) and 24 with hypothetical addressees, manipulated between subjects.

The display location of each picture in a quadruplet on any given trial was randomly determined. Quadruplets were presented in a fixed, randomly determined order such that displays that represented different conditions were distributed approximately evenly across the experiment, and constrained so that no more than one critical trial or three filler trials were presented consecutively.

Procedure. The task was administered with PsyScope 1.2.5 (Cohen et al., 1993). Speakers sat facing the computer screen. Addressees in the real-addressee condition sat beside speakers but facing the opposite direction, so that they could not see the computer screen. Each trial began with the message "⟨spacebar⟩," below which was displayed the trial number (to keep the speaker and addressee coordinated). 500 ms after the speaker pressed the space bar, the four pictures were displayed on the screen. 1000 ms after pictures-onset, an instruction appeared in the middle of the picture display that read either "2nd LAST" or "ALL." Speakers were instructed to describe the penultimate or all indicated pictures (in the same order as they were indicated) with each of these instructions, respectively. Three seconds after instruction onset (as the pictures remained on the screen), a small dot appeared just inside one of the four pictures for 200 ms. The dot then disappeared and immediately reappeared inside the next picture. In this manner, the dot indicated two, three, or all four of the displayed pictures. On the 18 critical trials, three pictures were indicated (a filler, then the target, and then the foil). On 18 of the filler trials, two pictures were indicated and on the other 18 of the filler trials, four pictures were indicated. Half of each of these sets of 18 pictures included simple fillers (as defined in above) and the other half included complex fillers. After the final dot disappeared, the pictures and instruction remained on the screen until the voice key detected that the speaker began his or her response. Upon voice-key detection, the screen cleared, and 3000 ms later the next "⟨spacebar⟩" instruction appeared.

Addressees, when present, were instructed to listen to speakers' descriptions, and indicate by writing "1" through "4" the pictures that speakers described in order. If unsure, addressees were instructed to guess, and to write a question mark beside their guess. Direct interaction between speaker and addressee was not explicitly encouraged and rarely occurred.

Each session began with instructions and a practice session. The practice session included five quadruplets, which were presented twice (to increase the amount of practice). Two practice trials had nonlinguistic-ambiguities and one had a linguistic-ambiguity. None of the pictures in the practice trials were used in the experiment proper.

Analyses. We transcribed all critical utterances, including all indications of disfluency. All trials on which speakers did not name the target object (e.g. they failed to respond, they said something like "I don't know," or they obviously never described the target picture) were discarded. This eliminated 17 utterances (2.1% of all critical trials). The remaining target descriptions were coded as *bare homophones* if the speaker described the target with an unmodified homophone label (i.e. the ambiguous label evaluated in the norming procedure, like *bat*; bare homophones modified only by indefinite specifiers like *a* and *some* were also coded as bare homophones). A target description was not coded as a bare homophone if it was a modified homophone (e.g. *small bat* or *flying bat*) or if speakers used an entirely different head noun (e.g. calling alphabetic letters *alphabet*). The first fluent attempt at a target description was coded (e.g. for *uh…smoking pipe*, *smoking pipe* was coded, but for *glasses…optical*, only *glasses* was coded). In addition, the category of *temporarily ambiguous* was coded. This category applied only to the *all* instruction condition, and was used when the speaker described the target object with a bare homophone, but then described the foil object without a bare homophone (e.g. *swan, bat, and baseball bat*). Such utterances are of interest because though the target description is ambiguous, the utterance taken as a whole is not. We discuss this issue further in the discussion.

For analyzable trials, we calculated the proportion of responses that were bare homophones per speaker and per item in each experimental condition. These proportions were submitted to three-way repeated-measures analyses of variance (ANOVAs) to yield analyses by speakers ($F1$) and items ($F2$). The ANOVA designs were $3\times2\times2$, with the factors ambiguity type (nonlinguistic, linguistic, control), instruction type (all, second last), and addressee type (real, hypothetical). By speakers, this design was mixed with ambiguity type and instruction type within speakers and addressee type between speakers, whereas by items, this was a fully repeated-measures design. Two speakers (both in the real-addressee condition) were inadvertently assigned to the wrong list and so were excluded. Variability is reported with 95% confidence-interval halfwidths based on single degree-of-freedom comparisons (Loftus & Masson, 1994). Pairs of means were compared either with simple main effects or pairwise comparisons, as appropriate. Significant effects reached a P level of .05 or less. All analyses were conducted also with arcsine-transformed proportions (Winer, Brown, & Michels, 1991), and similar results were observed. For readability, we report analyses conducted on raw proportions, and we report proportions as percentages.

1.2. Results

Fig. 3 shows the percentages of bare homophones that speakers produced as a function of ambiguity condition and instruction separately for speakers paired with real addressees (left graph) and hypothetical addressees (right graph). Speakers produced somewhat fewer bare homophones in the linguistic-ambiguity condition (middle bars: 40%) and many fewer bare homophones in the nonlinguistic-ambiguity condition (left bars: 1%), compared to in the unambiguous control condition (right bars: 65%). Furthermore, in the ambiguous conditions, speakers produced the same percentages of bare homophones regardless of whether the instruction was to describe all (hashed bars) or

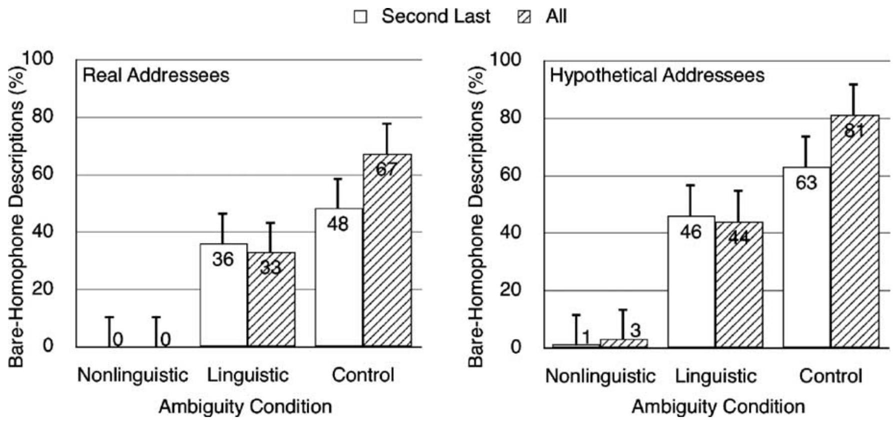

Fig. 3. Mean percentages of bare homophone descriptions to real and hypothetical addressees as a function of ambiguity type and instruction condition for Experiment 1.

just the second-last (open bars) indicated objects (1% in each case in the nonlinguistic-ambiguity condition, 39 vs. 41% in the linguistic-ambiguity condition). Unexpectedly, in the unambiguous control condition, speakers produced more bare homophones with the *all* instruction (74%) than with the *second-last* instruction (56%). Finally, although speakers produced fewer bare homophones overall when they addressed real addressees (left graph: 31%) compared to hypothetical addressees (right graph: 40%), they did so just as much in the unambiguous control condition (a 14% difference) as in the linguistic-ambiguity condition (an 11% difference). (Bare homophone production was unaffected by addressee type in the nonlinguistic-ambiguity condition, probably because speakers produced almost no bare homophones in either case).

These patterns were supported by statistical analyses. The main effect of ambiguity condition was significant, $F1(1,88)=122$, $\mathrm{CI}=\pm 8.1\%$; $F2(2,34)=71.2$, $\mathrm{CI}=\pm 11.1\%$. Pairwise comparisons revealed that speakers produced fewer bare homophones in the linguistic-ambiguity condition than in the control condition, $t1(88)=6.08$, $t2(34)=4.60$, and that they produced fewer bare homophones in the nonlinguistic-ambiguity condition than in the control condition, $t1(88)=15.5$, $t2(34)=11.8$. The main effect of instruction was marginally significant by speakers only, $F1(1,44)=4.03$, $P<.06$, $\mathrm{CI}=\pm 5.6\%$; $F2(1,17)=2.99$, $\mathrm{CI}=\pm 7.8\%$, and the instruction by ambiguity condition interaction was significant by speakers and items, $F1(1,88)=3.81$, $\mathrm{CI}=\pm 10.9\%$; $F2(1,34)=3.37$, $\mathrm{CI}=\pm 10.8\%$. Simple main effects revealed that the effect of instruction was not significant in the nonlinguistic- and linguistic-ambiguity conditions (all $F\mathrm{s}<1$), whereas the effect of instruction was significant in the control condition, $F1(1,88)=10.6$; $F2(1,34)=8.60$. Finally, the only significant effect of addressee type was a main effect, $F1(1,44)=5.04$, $\mathrm{CI}=\pm 8.1\%$; $F2(1,17)=12.4$, $\mathrm{CI}=\pm 5.1\%$. Simple main effects revealed that the difference between addressee types in the nonlinguistic-ambiguity condition was nonsignificant, $F\mathrm{s}<1$; the difference in the linguistic-ambiguity condition was significant by items only, $F1(1,88)=3.39$, $P<.08$, $F2(1,34)=4.61$; and the difference in the control condition was significant by speakers and items, $F1(1,88)=6.09$, $F2(1,34)=6.58$.

1.3. Discussion

Experiment 1 revealed four primary results of interest. First, just 1% of speakers' utterances were bare homophones in the nonlinguistic-ambiguity condition. This shows that speakers can effectively avoid nonlinguistic-ambiguity.

Second, 40% of speakers' utterances were bare homophones in the linguistic-ambiguity condition. Though this reflects a moderately high level of bare homophone production (compared to, say, the 1% level observed in the nonlinguistic-ambiguity condition), it was significantly lower than the 65% level observed in the unambiguous control condition. Thus speakers did reliably avoid linguistic ambiguities in this task. Nevertheless, the 40% level suggests that speakers may have failed to avoid linguistically ambiguous utterances relatively often; this issue is explored further below and in Experiments 2 and 3.

Third, when instructions required that speakers apply production processes to the foil object as well as to the target object (with the *all* instruction, compared to with the *second-last* instruction), they did not detect the linguistic-ambiguity of their expressions any better. This suggests that in Experiment 1, applying production processes to the ambiguity-causing foil (as well as to the target) did not improve linguistic-ambiguity detection. This implies that to the extent that speakers avoided linguistic-ambiguity at all under these conditions, they used comprehension-based processes rather than production-based processes.

An alternative explanation for the ineffectiveness of the instruction manipulation might be that speakers effectively performed the same task in the *all* and the *second-last* case. Speakers may have done so if they found it useful to covertly name all objects in either case, but then only overtly name the necessary objects once all objects were indicated. Two aspects of the procedure were designed to discourage this strategy, however. First, note that the instruction appeared before the indicating dots, meaning that speakers knew ahead of time whether they would need to formulate one description or multiple descriptions. If speakers consider it harder to formulate multiple descriptions, they should not formulate descriptions of all indicated objects in the *second-last* condition. Second, note that the indicating dots appeared for just 200 ms per object; this made it impossible to formulate object names as they were indicated, requiring speakers to do so instead after the objects were indicated. But after the objects were indicated, speakers could *overtly* describe the objects, making a strategy of covertly naming all objects and then overtly naming just the second-last object markedly more effortful than simply just naming the objects they needed to describe. Furthermore, if speakers effectively performed the same task in both instruction conditions, we would expect they would take equally long to formulate their utterances in either case. Instead, speakers took significantly longer to begin their utterances in the *all* condition (2534 ms) than in the *second-last* condition (2120 ms), paired $t(44)=4.44$ (this analysis included only reaction time between 500 and 5000 ms, and collapsed across all other independent variables; one subject was eliminated due to missing values). This suggests that speakers engaged in more formulation before beginning articulation in the *all* than in the *second-last* condition, arguing against the possibility that they applied production processes before description to the same extent in both cases.

It is worth noting that with the *all* instruction, of the 53 bare homophones that speakers produced in the linguistic-ambiguity condition, 43 occurred in temporarily ambiguous utterances like *swan, bat, and baseball bat*. In temporarily ambiguous utterances, the target descriptions (*bat*) are ambiguous, but foil descriptions (*baseball bat*) are not. Consequently, an addressee could in principle determine (via deduction) that the target description applies to its intended referent. This raises the possibility that when speakers produced such utterances, they might have detected that they were about to produce ambiguous target descriptions, but they articulated those ambiguous target descriptions anyway, because they recognized they could subsequently disambiguate the foil and therefore produce an utterance that was unambiguous as a whole. However, speakers almost never produced temporarily ambiguous utterances in the nonlinguistic-ambiguity condition (just once, producing *chopsticks, ring, and bigger ring*), and instead produced unambiguous target descriptions almost without fail (136 times, as in *chopsticks, smaller ring, bigger ring*). The fact that speakers nearly always disambiguated target descriptions in the nonlinguistic-ambiguity condition suggests that when speakers can determine that a target description will be ambiguous, they disambiguate it. In turn, this implies that when speakers used temporarily ambiguous utterances, they did not know that the target descriptions were ambiguous before (or as) they produced them.

Unexpectedly, in the control condition, speakers produced more bare homophones with the *all* versus with the *second-last* instruction. This difference is unlikely to be due to overall differences in verbosity, as target descriptions were about equally long in each instruction condition (at least as measured by the length of the transcriptions). One explanation for the difference is that speakers may have adopted a general strategy of avoiding ambiguity by describing objects overspecifically, unless they detected the absence of ambiguity, and so they could use a simpler (bare homophone) label. If so, then when speakers described all indicated pictures, they may have been better able to detect the absence of ambiguity (and so they did not need an overspecific description), compared to when they described just the second-last indicated picture. This possibility is especially surprising in light of the fact that speakers were unaffected by instruction in the linguistic-ambiguity condition, meaning that this interpretation implies that the *all* instruction may have allowed speakers to better detect the absence of ambiguity in the control condition at the same time as it did not allow speakers to better detect the presence of ambiguity in the linguistic-ambiguity condition. We leave further exploration of this unanticipated result to future research.

The fourth result of interest is that speakers did not avoid ambiguity any better when they addressed real addressees, compared to when they addressed hypothetical addressees. Speakers produced fewer bare homophones overall when they addressed real addressees, doing so just as much when they described unambiguous displays (in the control condition) as when they described linguistically ambiguous displays. This shows that with real addressees, speakers simply described targets more specifically overall, and therefore produced fewer bare homophones across both ambiguous and unambiguous conditions. In fact, this may constitute a nonspecific strategy that speakers use to avoid ambiguity under tricky communication situations—by describing objects more specifically than usual, they may be more likely to provide distinguishing information that circumvents potential ambiguity.

In sum, Experiment 1 showed that speakers avoided nonlinguistic-ambiguity to a near-total extent and that they avoided linguistic-ambiguity to at least some extent, doing so in the latter case with a comprehension-based rather than a production-based ambiguity-detection strategy. However, as noted above, speakers still produced bare homophones 40% of the time in the linguistic-ambiguity condition. Might this mean that speakers did not avoid linguistic-ambiguity especially effectively in Experiment 1?

In fact, the temporarily ambiguous utterances that speakers commonly produced with the *all* instruction are consistent with this possibility. As noted above, the fact that speakers often called a flying mammal *bat* in utterances like *swan, bat, and baseball bat* implies that they failed to detect that they prepared a target description that could refer not only to the target object, but also to the about-to-be-described foil. However, the fact that speakers subsequently called a baseball bat *baseball bat* suggests that they may have successfully recognized that a bare homophone label like *bat* could refer not just to the foil object, but also to the just-described target object. If so, then speakers' temporarily ambiguous utterances might show that they can detect linguistic-ambiguity better than the 40% level observed in Experiment 1 indicates, but only *after* they already articulated the ambiguous description of one of the potentially ambiguous referents. Interestingly, this analysis of temporarily ambiguous utterances follows naturally from the production-based ambiguity detection strategy outlined above, as it implies that when speakers retrieve the same term for two distinct meanings, they can detect the (linguistic) ambiguity of that term. The fact that speakers produced temporarily ambiguous rather than wholly unambiguous utterances adds to this strategy the restriction that it might only apply after-the-fact—after speakers have retrieved and articulated one ambiguous label.

That said, the present data do not uniquely support this explanation. Most obvious is that the temporarily ambiguous utterances may have resulted simply from idiosyncratic properties of the foils (e.g. perhaps speakers simply tend to call a baseball bat *baseball bat*). Experiments 2 and 3 sought to determine whether speakers really can use a production-based strategy to detect ambiguity after-the-fact, by having speakers describe the same target objects either before they described foils (as with Experiment 1's *all* instruction), or after they described foils. If this explanation is correct, then speakers ought to use fewer bare homophones when they describe targets after they describe foils, compared to when they describe targets before foils, and they should produce fewer bare homophones than the 40% level observed in Experiment 1.

2. Experiments 2 and 3

In Experiments 2 and 3, speakers saw displays like in Experiment 1, and we manipulated whether the dot that indicated the foil appeared before or after the dot that indicated the target. On critical trials, all four objects were indicated, so that the foil object would not be described first. Because Experiment 1 revealed that speakers did not avoid ambiguity per se any more effectively with real addressees, all speakers in Experiments 2 and 3 addressed hypothetical addressees. Experiment 3 was highly similar to Experiment 2, the primary difference being that in Experiment 2, speakers described pictures in the same order as the dots appeared, whereas in Experiment 3, they described pictures in the opposite order. We carried out Experiment 3 to ensure that performance was not due to overall ease of the task (as describing pictures in the opposite order as they are indicated is much harder than describing them in the same order).

2.1. Method

Speakers. Each experiment included 48 different speakers from the same population as Experiment 1.

Apparatus. Apparatus was the same as in Experiment 1.

Materials, design, and procedure. Materials, design, and procedure were the same as in Experiment 1, except for four changes in Experiment 2: (a) On critical trials, the dots indicated all four pictures instead of just three, and the filler trials from Experiment 1 on which four pictures were indicated were modified so that three pictures were indicated. (b) The instruction manipulation was dropped, and speakers described all indicated pictures on every trial. Instead, we manipulated *foil position*: Either the foil was indicated (and thus was to be described) before the target (*foil-then-target*) or after the target (*target-then-foil*). The target picture was always indicated third, so that the foil picture was indicated second in the foil-before condition and was indicated fourth in the foil-after condition. (c) Because speakers were to describe all indicated pictures, no trial-by-trial instruction was necessary. We instead instructed speakers at the beginning of the experiment to always describe all indicated pictures. The event sequence on each trial was as in Experiment 1, except that the dots began appearing three seconds after pictures onset. (d) Pilot work indicated that speakers had difficulty maintaining information about all four pictures in memory during description, so in Experiment 2, pictures remained on the screen, and speakers pressed the space bar at the end of their descriptions to advance to the next trial. Experiment 3 included three additional changes of its own: (a) The order that pictures were indicated in Experiment 3 was the opposite of that in Experiment 2. (b) Speakers were instructed to describe the pictures in the opposite order as they were indicated. (c) After the final dot disappeared from the screen, the pictures remained for 1000 ms plus an additional 500 ms for each picture that was to be described.

Analyses. Utterances were transcribed and coded as in Experiment 1. We additionally eliminated any trial where the speaker either obviously did not describe the foil (i.e. just as we did for targets), or where the order of description was incorrect in a way that compromised the order manipulation. This eliminated 9 (1.0%) and 98 (11.3%) trials in Experiments 2 and 3, respectively. Two speakers were excluded from the speakers analysis of Experiment 3 because dropped trials led to an empty cell in each design. Note that the greater exclusions in Experiment 3 suggest that its procedure was indeed more difficult than that of Experiment 2. Data were analyzed as in Experiment 1, except that the addressee type factor was dropped and the foil position factor replaced the instruction factor.

2.2. Results

The mean percentages of bare homophones that speakers produced as a function of ambiguity type and foil position are shown in Fig. 4 for Experiment 2 and in Fig. 5 for Experiment 3. Overall, speakers in Experiment 2, like those in Experiment 1, produced

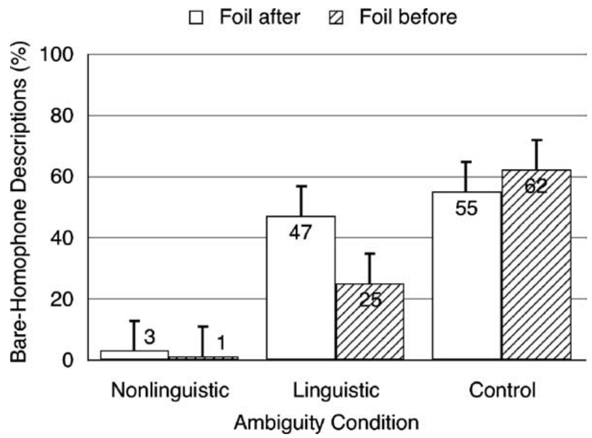

Fig. 4. Mean percentages of bare homophone descriptions as a function of ambiguity type and foil position for Experiment 2 (pictures were described in the same order as they were indicated).

bare homophones somewhat less (36%) in the linguistic-ambiguity condition and much less (2%) in the nonlinguistic-ambiguity condition, compared to in the control condition (59%). Most importantly, foil position affected bare homophone production only in the linguistic-ambiguity condition, as speakers produced 22% fewer bare homophones when they described foils before targets (25%), compared to after (47%). These observations were supported by statistical analyses. The main effect of ambiguity condition was significant, $F1(2,94)=102$, $\mathrm{CI}=\pm 7.9\%$, $F2(2,34)=45.5$, $\mathrm{CI}=\pm 12.2\%$. Pairwise comparisons revealed that speakers produced significantly fewer bare homophones in the linguistic-ambiguity condition, $t1(94)=5.7$, $t2(34)=3.8$, and the nonlinguistic-ambiguity condition, $t1(94)=14.2$, $t2(34)=9.5$, compared to in the control condition.

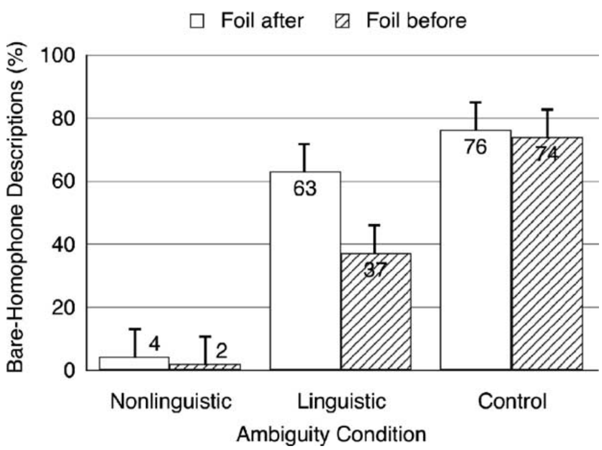

Fig. 5. Mean percentages of bare homophone descriptions as a function of ambiguity type and foil position for Experiment 3 (pictures were described in the opposite order that they were indicated).

The interaction between ambiguity condition and foil position was also significant, $F1(2,94)=9.13$, CI $= \pm 9.7\%$, $F2(2,34)=10.7$, CI $= \pm 8.8\%$. The simple main-effect of foil position was significant only within the linguistic-ambiguity condition, $F1(1,94)=19.6$, $F2(1,34)=23.3$.

Experiment 3 showed the same pattern as Experiment 2 except more extremely. Overall, speakers produced somewhat fewer (50%) and many fewer (3%) bare homophones in the linguistic- and nonlinguistic-ambiguity conditions, respectively, compared to in the control condition (75%). Speakers also produced 26% fewer bare homophones when the foil preceded the target (37%) compared to when it followed the target (63%). This pattern led to a significant main effect of ambiguity condition, $F1(2,90)=119$, CI $= \pm 9.4\%$, $F2(2,34)=111$, CI $= \pm 10.1\%$. Pairwise comparisons revealed that both the linguistic and the nonlinguistic-ambiguity conditions were produced with significantly fewer homophones than the control condition [linguistic: $t1(90)=5.3$, $t2(34)=4.9$; nonlinguistic: $t1(90)=15.2$, $t2(34)=14.7$]. Ambiguity condition and foil position interacted, $F1(2,90)=8.94$, CI $= \pm 9.2\%$, $F2(2,34)=7.52$, CI $= \pm 11.8\%$. The simple main effect of foil position was again significant only in the linguistic-ambiguity condition, $F1(1,90)=33.2$, $F2(1,34)=24.0$.

2.3. Discussion

Experiments 2 and 3 showed that speakers described targets with bare homophones significantly less often when they had already described foils, compared to when they had not yet described foils (even though they were just about to). This suggests that the comprehension-based ambiguity-detection strategy speakers used in Experiment 1 was not fully effective, and that an after-the-fact production-based ambiguity-detection strategy was notably more so. In fact, comparing Experiment 2 to Experiment 1 reveals that when speakers described targets before foils in Experiment 2, they produced about as many bare homophones as they did in Experiment 1 with the *all* instruction (with hypothetical addressees; 47% in Experiment 2 vs. 46% in Experiment 1). Against this level, when speakers described targets after foils, they produced over 20% fewer bare homophones (25%). Thus, the different levels of bare homophone production within the linguistic-ambiguity condition of Experiment 2 reveals the extent to which a production-based ambiguity detection strategy can better detect ambiguity after it is produced than a comprehension-based one can before it is produced. (Note that the large task differences between Experiments 1 and 3 make analogous comparisons between the two difficult.) It is further worth noting that in Experiment 2, when speakers described targets before foils, they did not use bare homophones significantly less often in the linguistic-ambiguity condition than in the unambiguous control condition (an 8% difference as compared to a 9.7% confidence interval). This reinforces the claim that the comprehension-based strategy is not highly effective at detecting ambiguities before they are produced.

When speakers described targets after foils in the linguistic-ambiguity condition, was the 25% level of bare homophone production *completely* unambiguous? To assess this, we conducted a study in which an independent group of speakers saw linguistically ambiguous displays (like Fig. 2b but without the dots) for 5 s, and then were asked to unambiguously describe target objects. Presumably, with a 5-s preview, speakers could design optimal descriptions for the objects in the display. Results showed that 30% of speakers' utterances in such optimal descriptions were bare homophones. This suggests that the 25% level observed when speakers described targets after foils in Experiment 2 was close to completely unambiguous.

3. General discussion

The experiments presented here revealed three primary results. First, across all experiments and conditions, speakers almost never described targets with bare homophones in nonlinguistically ambiguous displays, showing that they effectively avoided nonlinguistic-ambiguity. Second, speakers in Experiment 1 described targets with bare homophones less often in linguistically ambiguous displays compared to in control displays, showing that they can to some extent detect and avoid linguistically ambiguous expressions before producing them. The fact that ambiguity detection was no better with the *all* instruction, when task demands required that speakers name the foil as well as the target, suggests that to the extent that speakers detected ambiguity before producing it, they used a comprehension-based rather than production-based strategy. Third, speakers in Experiments 2 and 3 described targets with bare homophones less often after they described the ambiguity-causing foil compared to before, showing that they can detect ambiguities in their expressions to an even greater extent after producing them. Such after-the-fact ambiguity detection follows from a production-based mechanism that determines whether speakers retrieved the same word-form for multiple distinct meanings.

These experiments imply that it is important to distinguish nonlinguistic- and linguistic-ambiguity avoidance. Whereas nonlinguistic-ambiguities are effectively avoided, the situation with linguistic ambiguities is more complex. Specifically, Experiment 1 vs. Experiments 2 and 3 illustrate a revealingly complementary pattern of linguistic-ambiguity avoidance. Before producing an ambiguous expression, speakers evidently have available only a comprehension-based ambiguity-detection strategy: Without the assistance of production processes, speakers sometimes can comprehend a to-be-produced linguistic form and determine whether it means more than one thing in the present context. However, this comprehension-based strategy is not especially effective, as it detects some but not all ambiguous expressions before speakers produce them. In complement, after speakers produce a linguistically ambiguous expression, production-based processes can better detect that linguistic-ambiguity. The greater effectiveness of production-based ambiguity detection (even if after-the-fact) is important to recognize for three reasons. First, it illustrates that the comprehension-based strategy that speakers use to detect ambiguous expressions before producing them is suboptimal, in comparison to the production-based strategy that speakers use after ambiguous production. Second, it illustrates that speakers are not oblivious or unconcerned with the ambiguity of their utterances. The fact that the same target objects were described less ambiguously when ambiguity-causing foils were described beforehand shows that when speakers detect that an expression will be ambiguous, they avoid it. Third, an after-the-fact ambiguity-avoidance strategy could be important for naturalistic communication, as it implies that speakers might clarify or correct already articulated descriptions.

Given that production processes can so effectively detect linguistically ambiguous expressions after they are produced, why do those production processes not operate to detect linguistically ambiguous expressions before they are produced? Note that for production processes to detect an ambiguous expression before it is produced, speakers must retrieve two descriptions—for the target and for the foil—before the first is articulated. If production at the level of word-forms is incremental—if speakers articulate words as they are retrieved—then the second description will not be retrieved in time for speakers to modify the first description (see Griffin, 2001, for evidence of the limited scope of lexical planning in these kinds of utterances, but see Freedman, Martin, and Biegler, 2004). However, note that this same incrementality means that as the second description is formulated, the form of the first description can be available to speakers, either via perception or production memory. In short, the incrementality of the language-production process may both prevent the more effective production-based ambiguity-detection strategy from detecting ambiguous expressions before they are produced and enable that strategy to detect ambiguous expressions after they are produced.

Why can't a comprehension-based ambiguity detection strategy operate more effectively? It is often claimed that comprehension monitoring is a cognitively complex task (see Levelt, 1989), such that it can sometimes fail to detect production problems. For example, speakers sometimes fail to detect their speech errors, especially when production is difficult (e.g. Oomen & Postma, 2002). Note that monitoring for ambiguity is likely to be even more difficult than monitoring for errors, because only ambiguity monitoring requires that speakers reject a to-be-produced utterance even though it actually matches their intended meaning.

Taken to their conclusion, these results raise the worry that because speakers often fail to avoid linguistically ambiguous expressions before they are articulated, communication itself may be compromised. However, as noted above, that speakers can effectively detect ambiguity after it is produced could be useful in natural production situations. Furthermore, comprehenders may take on much of the burden for dealing with linguistic-ambiguity. For example, the threat posed by linguistic-ambiguity may be reduced if comprehenders are sensitive to a wide range of factors that dictate speakers' particular usages (see Roland, Elman, & Ferreira, 2003), or if comprehenders use pragmatic factors to successfully guess speakers' intended meanings (in the context of a flying mammal and a baseball bat, comprehenders can figure out what an instruction to "pick up the bat" should mean). Put somewhat differently, the present results show not that communication is unsuccessful, but that communication is successful despite speakers' (first-pass) insensitivity to the linguistic-ambiguity of their own expressions.

Moreover, speakers' insensitivity to linguistic-ambiguity constitutes an important consideration in linguistic and psycholinguistic theories. It is commonly claimed that speakers choose particular utterances to avoid linguistic-ambiguity (see Temperley, 2003, for a recent example). The present analysis suggests that at least as far as online production is concerned, linguistic-ambiguity may not powerfully influence speakers' utterances, at least not until after they have been articulated, and regardless, communication may be just as successful. Of course, it is possible that linguistic-ambiguity can influence production offline, by changing the grammatical possibilities that the language offers (perhaps through the language-acquisition process). This highlights that it is important to distinguish the possibility that ambiguity affects *speakers* from the possibility that ambiguity affects *grammars*. These results constrain the former kind of influence.

It is also worth noting that in these experiments (as in others), speakers avoided nonlinguistic-ambiguity almost without fail. According to the model shown in Fig. 1, this suggests that speakers are highly sensitive to the possibility that multiple referents in a communicative situation are similar in meaning. Of course, this need not have been the case; similarity of meaning, unlike animacy or agency, is not obviously a grammatically relevant feature of meaning. Speakers are sensitive to similarity anyway probably because nonlinguistic-ambiguity, unlike linguistic-ambiguity, is likely to be both quite frequent and quite disruptive. Nonlinguistic-ambiguity can occur in nearly any situation where multiple tokens of a common type co-occur. As long as people tend to collect in groups, cars in parking lots, and trees in forests, nonlinguistic-ambiguity is a threat. And whereas the difference in meaning between linguistically ambiguous referents might allow comprehenders to pragmatically reason past a linguistic-ambiguity, the similarity in meaning between nonlinguistically ambiguous referents probably prevents comprehenders from pragmatically reasoning past a nonlinguistic-ambiguity (e.g. a direction to "pick up the bat" is highly ambiguous in the face of multiple baseball bats).

From an information-processing perspective, the effectiveness of speakers' after-the-fact linguistic disambiguation raises important issues. Specifically, how do speakers detect that they have retrieved two identical word forms? One possibility, raised in the introduction, is that speakers monitor their productions to determine whether a to-be-used form is the same as a form that was previously used for a different meaning. A different and intriguing possibility is that speakers detect the use of two identical word forms more implicitly. Specifically, it may be that when speakers produce a linguistic form, its superficial representation enters a kind of refractory period, such that speakers find it more difficult to immediately repeat that form (see Wheeldon, 2003, for relevant evidence and Dell, 1986, for a theoretical motivation for such refractory periods). Any such difficulty of superficial repetition may implicitly act as a processing signal that informs speakers that they are about to produce the same form twice, which might trigger the search for a distinguishing description. This implicit mechanism has implications for theories of word production. First, it raises difficulties for theories that claim that homophones have distinct representations at the level where lexical items are selected (Levelt et al., 1999), because speakers would not then attempt to select the same representation twice at that level. Instead, approaches that tie lexical selection more closely to phonological properties (including homophony) fare better (e.g. Cutting & Ferreira, 1999; Dell & Reich, 1981; Ferreira & Griffin, 2003; Ferreira & Humphreys, 2001). Second, for obvious reasons, this mechanism raises difficulties for theories that claim that homophones are not represented by a single representation anywhere in production (Caramazza, Costa, Miozzo, & Bi, 2001). In sum, these observations raise the possibility that after-the-fact ambiguity avoidance can be used to tease apart competing claims of how word production operates.

Overall, the present results show that when it comes to avoiding ambiguity, speakers are only somewhat cooperative because they adopt the perspective of their listeners, by comprehending their utterances before producing them. Speakers can be yet more cooperative—avoid ambiguity better—because of the action of production processes themselves, though only after-the-fact. Of course, communication ultimately succeeds. The present results indicate that this is so because speakers (and their grammars) balance the threat of ambiguity against the information-processing demands required for detecting and avoiding it.

Acknowledgements

Portions of this work were presented at the Architectures and Mechanisms for Language Processing conference in August, 2003. This research was supported by National Institute of Health grant R01 MH-64733. We thank Boaz Keysar, Neal Pearlmutter, and the members of the psycholinguistics groups at UCSD for helpful discussions, Gary Dell, Zenzi Griffin, Martin Pickering (who pointed out the implicit avoidance strategy), and three anonymous reviewers for valuable comments on the manuscript, and Tom Christensen, Alma Correa, Cassie Gipson, Amy Gonzales, and Mary Tibbs for assistance collecting data. Please address correspondence to Vic Ferreira at the Department of Psychology 0109, University of California, San Diego, La Jolla, CA, 92093-0109, USA. Email: vferreira@ucsd.edu

Appendix

Table A1
Description of targets and linguistically ambiguous foils used in all experiments

Target object	Linguistically ambiguous foil
Bat-animal	Bat-baseball
Bow-arrow	Bow-ribbon
Chest-torso	Chest-treasure
Diamond-gem	Diamond-shape
Fan-cool	Fan-sports
Glasses-spectacles	Glasses-drink
Heart-shape	Heart-organ
Letters-alphabet	Letters-mail
Mouse-computer	Mouse-animal
Nail-hammer	Nail-finger
Nuts-food	Nuts-bolts
Pi-3.14	Pie-food
Pipe-smoke	Pipe-water
Plug-sink	Plug-electric
Ring-engagement	Ring-boxing
Slides-playground	Slides-projector
Tank-army	Tank-aquarium
Tape-adhesive	Tape-cassette

References

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 714–735.

Bates, E., D'Amico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., et al. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin and Review*, *10*(2), 344–380.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*, 1–47.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1430–1450.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, and Computers*, *25*, 257–271.

Craig, T. B., Nicol, J. L., & Barss, A. (1995). *Failure to avoid syntactic ambiguity during sentence production: An analysis of spontaneous speech corpora*. Paper presented at the eighth annual CUNY conference on human sentence processing, Tucson, AZ.

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 318–344.

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

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*, 611–629.

Elsness, J. (1984). That or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies*, *65*, 519–533.

Ferreira, V. S., & Cutting, J. C. (1997). Ninety-three pictures and 108 questions for the elicitation of homophones. *Behavior Research Methods, Instruments and Computers*, *29*(4), 619–635.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340.

Ferreira, V. S., & Griffin, Z. M. (2003). Phonological influences on lexical (mis)selection. *Psychological Science*, *14*, 86–90.

Ferreira, V. S., & Humphreys, K. R. (2001). Syntactic influences on lexical and morphological processing in language production. *Journal of Memory and Language*, *44*(1), 52–80.

Freedman, M. L., Martin, R. C., & Biegler, K. (2004). Semantic relatedness effects in conjoined noun phrase production: Implications for the role of short-term memory. *Cognitive Neuropsychology, 21, 245–265.*

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower, *The psychology of learning and motivation* (Vol. 9) (pp. 133–177). New York: Academic Press.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*(1), B1–B14.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, *42*(2), 113–157.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117.

Kraljic, T., & Brennan, S. E., (2003). *Prosodic disambiguation of syntactic structure: For the speaker or for the addressee?* Paper presented at the CUNY conference on human sentence processing, Boston, MA.

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

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–75.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin and Review*, *9*(3), 550–557.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review*, *1*, 476–490.

Mims, K., & Trueswell, J. C. (1999). *Do speakers help listeners? Prosodic cues, lexical cues, and ambiguity avoidance*. Poster presented at the 12th annual CUNY conference on human sentence processing, New York, NY.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336.

Oomen, C. C. E., & Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes*, *17*(2), 163–184.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, *77*, 97–131.

Roland, D. W., Elman, J. L., & Ferreira, V. S. (2003). *Signposts along the garden path*. Paper presented at the 16th annual CUNY conference on human sentence processing, Boston, MA.

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, *29*(2), 169–182.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*(1), 1–24.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103–130.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

Temperley, D. (2003). Ambiguity avoidance in English relative clauses. *Language*, *79*, 464–484.

Wardlow, L., & Ferreira, V. S. (2003). *Finding common ground: How do speakers block privileged information?* Paper presented at the AMLaP 2003, Glasgow, Scotland.

Wheeldon, L. (2003). Inhibitory form priming of spoken word production. *Language and Cognitive Processes*, *18*(1), 81–109.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed). New York: McGraw-Hill.