Anchoring Comprehension in Linguistic Precedents

Dale J. Barr

Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana–Champaign

and

Boaz Keysar

The University of Chicago

Past research has shown that when speakers refer to the same referent multiple times, they tend to standardize their descriptions by establishing linguistic precedents. In three experiments, we show that listeners reduce uncertainty in comprehension by taking advantage of these precedents. We tracked listeners' eye movements in a referential communication task and found that listeners identified referents more quickly when specific precedents existed than when there were none. Furthermore, we found that listeners expected speakers to adhere to precedents even in contexts where it would lead to referential overspecification. Finally, we provide evidence that the benefits of linguistic precedents are independent of mutual knowledge—listeners were not more likely to benefit from precedents when they were mutually known than when they were not. We conclude that listeners use precedents simply because they are available, not because they are mutually known. © 2001 Elsevier Science (USA)

Key Words: psycholinguistics; mutual knowledge; comprehension; precedents; reference.

In Graham Greene's novel *The End of the Affair*, the narrator Bendrix relates the aftermath of his tryst with Sarah, the wife of his friend Henry. Sarah had ended her affair with Bendrix for mysterious reasons, which caused him to suspect that she had started seeing someone else. Bendrix's jealousy compelled him to secretly hire a private detective to spy on her. One day, Bendrix invites Henry to lunch to tell him about some reports that the detective had provided. Bendrix avoids the topic over lunch, chatting instead about Henry's work in the British Ministry of Pensions. Recently, Henry had been involved in a Royal Commission that the British press reported on daily, making him something of a public figure. After the meal, Bendrix finally discloses to Henry that he had hired a detective to follow Sarah and tells him, "I think you ought to read the reports." Just then, a friend of Bendrix's approaches to say hello, and Bendrix introduces the newcomer to Henry. The newcomer tells Henry, "I've been following the reports every day." Henry is confused and asks, "What reports?" Bendrix's friend replies, "The Royal Commission." The narrator observes that, "for once, Henry's work had not come first to mind when that word was uttered."

When the newcomer said "the reports," why did Henry think first of the reports on his wife instead of the press reports? In linguistic terms, Henry's behavior was egocentric—he interpreted "the reports" according to Bendrix's use of the term, which caused him to think of the reports on his wife. He ignored the fact that referential meaning depends on mutual knowledge, or the set of beliefs and assumptions that people share and know that they share (Clark & Marshall, 1981). Henry had little reason to assume that the newcomer would be using the term in the same way that Bendrix had because this earlier linguistic experience was not, in the terms of Clark and Marshall, "co-present" between him-

self and the newcomer. Instead, Henry should have realized that the only reports that were known to the two of them were the press reports on the Royal Commission.

We propose that Henry's misunderstanding reflects a normal feature of language processing: the egocentric anchoring of comprehension in linguistic precedents. When Bendrix uses the word "reports" to refer to the documents prepared by the detective agency, he establishes a precedent whereby the word "reports" designates a particular set of documents. Over the span of the conversation, the precedent serves as a linguistic index to a representation of the referent in memory.

While precedents might occasionally cause misunderstandings like Henry's, we propose that they generally enhance comprehension because they reduce variability in the content and the meaning of speech. When specific precedents exist, listeners can recognize words and identify referents more swiftly than when they are absent. Although precedents are established by specific speakers for specific purposes, we suggest that, like Henry, listeners use precedents because they are egocentrically available, not because they are mutually known.

Three experiments provide support for these proposals. We used eyetracking techniques in a "communication game" in which listeners interpreted speakers' referential expressions. Experiment 1 establishes the basic benefit of precedents to comprehension and examines the interplay of precedents with preexisting, conventional names. Experiments 2 and 3 test whether the precedent must be mutually known to obtain this benefit.

The concept of mutual knowledge, or common ground, has a long and controversial history, from early debate regarding its plausibility as a psychological mechanism (Clark, 1982; Clark & Carlson, 1982; Clark & Marshall, 1981; Sperber & Wilson, 1982) to its later widespread acceptance as an essential element in a general theory of language use. According to the prevailing view, mutual knowledge is a kind of "model of the interlocutor" that language users derive through the application of a set of copresence heuristics, which they use when they process utterances (Clark & Carlson, 1981; Clark & Marshall, 1981). From the standpoint of language production, Clark and Murphy (1982) suggested that speakers tailor utterances to their mutual knowledge with specific addressees. Clark and Carlson (1981) presented a similar proposal for language comprehension when they claimed that listeners should restrict the information they consider to mutual knowledge, or common ground: " . . . when a listener tries to understand what a speaker means, the process he goes through can limit memory access to information that is common ground between the speaker and his addressees" (p. 328). They suggested that comprehension process will be optimal, "if it limits its access to that common ground" (p. 328). Both proposals characterize mutual knowledge as something that speakers and listeners will spontaneously and routinely consider when they process utterances. Furthermore, they characterize language processing systems as optimal to the extent that these systems rely exclusively on mutually known information.

The Perspective Adjustment model of Keysar and colleagues opposes a strongly contrasting conception of optimality to the normative view of Clark and colleagues (Keysar & Barr, in press; Keysar, Barr, & Horton, 1998). According to this conception, the design of a language processing system is "optimal" not because it guarantees mutual understanding, but because it provides adequate real-time understanding at a minimal cognitive cost. Perspective Adjustment posits that language processing systems meet this standard by employing a "egocentric anchoring and adjustment heuristic." The model assumes a rapid, automatic egocentric anchoring process coupled with a slower, optional adjustment process that computes information about perspective. These processes need not operate strictly in serial fashion, but can function in parallel or in cascade. Perspective Adjustment assumes that the egocentric heuristic often is sufficient for mutual understanding. Hence, language users need not routinely check mutual knowledge every time they process an utterance. This view predicts that the kinds of errors a person will make in conversation will be systematically biased toward that person's own knowl-

edge. By anchoring egocentrically and then using mutual knowledge as needed to adjust to the interlocutor's perspective, language users trade-off failsafe understanding for cognitive efficiency (Keysar & Barr, in press; Keysar, Barr, & Horton, 1998). In recent work, Keysar and colleagues have found evidence supporting this model for both language production (Horton & Keysar, 1996) and language comprehension (Keysar, Barr, Balin, & Brauner, 2000; Keysar, Barr, Balin, & Paek, 1998).

Our interest in this article is to test these two views of the role of mutual knowledge with respect to linguistic precedents. We ask, When people understand referential expressions, do they interpret them egocentrically or against the background of mutual knowledge?

One good reason for listeners to use linguistic precedents is the fact that *speakers* use them. When speakers refer to the same referent multiple times in the same conversation, they tend to standardize their descriptions. Later descriptions tend to be shorter and less lexically diverse than early ones (Krauss & Weinheimer, 1964, 1966). When the members of a conversational dyad take turns describing a referent, they come to describe the referent in the same way. It has been proposed that descriptions can become standardized through common ground (Clark & Wilkes-Gibbs, 1986) or through language users' attempts to establish consistency between language production and comprehension systems (Garrod & Anderson, 1987).

In a recent proposal, Brennan and Clark (1996) argued that interlocutors coordinate reference by forming partner-specific "conceptual pacts," or agreements concerning how a referent is to be conceptualized. These conceptual-level agreements affect how referents are encoded linguistically. For instance, dyads formed pacts to refer to a particular shoe as *the loafer* when it was presented in the context of another shoe (a sneaker). Later, speakers continued to call it the loafer, even without the constraining context of the sneaker (i.e., when the basic-level term "shoe" would have been sufficient). They found that when speakers switched to a new addressee, they eventually stopped using the overspecified term they had established with the old partner.

They claimed that such conceptual pacts are "partner specific" in the sense that they are the end product of interaction between the speaker and listener: "the references in our task emerged not from solitary choices on the part of the director, but from an interactive process by both director and matcher" (p. 1491).

For our current purposes, the issue of interactivity can be viewed as orthogonal to that of precedent use. People can use precedents when speaking or understanding either because the precedents were interactively established or just because prior use has made them available. We seek a general theory of precedent use that covers both interactive and noninteractive situations. In this article, we use the term "linguistic precedents" to broadly characterize the word-referent mappings that listeners establish while comprehending discourse. The precedents that we investigate are established interactively in Experiments 1 and 2, but noninteractively in Experiment 3. In addition, although a speaker's initial naming of a referent involves classification (Brown, 1958), we remain agnostic regarding whether the precedent that is the result of this initial naming is anything other than lexical. For instance, if Bendrix had referred to the reports using the word "documents" instead of "reports," it seems that Henry's misunderstanding would have been much less likely, even though "documents" and "reports" entail quite similar conceptualizations. The issue of the level at which precedents persist is beyond the scope of our experiments.

Precedents such as conceptual pacts have been examined only as they pertain to language production. Our question is, Do listeners expect speakers to adhere to linguistic precedents? We propose that precedents could potentially benefit referential understanding by reducing uncertainty in speakers' intended meanings, especially when preexisting, conventional names are not readily available. The first experiment establishes this basic benefit of linguistic precedents to comprehension.

EXPERIMENT 1

We used a referential communication task in which one participant (the "director") instructed a second (the "addressee") to rearrange objects

in a grid (see Fig. 1) to match a picture. We tracked the position of addressees' eyes as they followed directors' instructions (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). We compared the length of time it took them to find previously mentioned referents versus referents that were previously unmentioned. We expected that addressees would take advantage of linguistic precedents to facilitate identification of referents.

As an example, consider an exchange about the set of objects located in the grid in Fig. 1. (For clarity of exposition, the figure contains fewer objects than a typical item in the experiment.) The set includes a folded piece of paper in the shape of an upside-down V (top right corner). When a speaker first refers to this object as the "tent," the addressee would need to establish that the folded paper is the object that could most plausibly be conceptualized as a "tent." By accepting the label, then, the addressee establishes a precedent by which this object is labeled "tent." In future instances when the speaker uses the label "tent," the addressee could take advantage of the precedent to quickly identify the referent.

Yet this benefit might be due to factors other than the establishment of a linguistic precedent. Specifically, addressees might be faster in finding previously mentioned objects simply because they learned the location of the objects in the grid. To control for this, we also manipulated whether referents had or lacked preexisting conventional names ("conventional" versus "unconventional" referents). Conventional referents were familiar objects that have widely known conventional names, like the apple in Fig. 1. We would expect near uniform agreement among speakers on how they would name these referents. Unconventional referents were objects that lacked these common names, like the folded paper in Fig. 1. Unlike conventional referents, one would expect wide variation in how speakers would refer to them.

Objects such as the apple are less likely to benefit from a specific naming precedent compared to unconventional referents. In contrast, knowledge of a referent's location should contribute equally to identification speed for both kinds of referents. Therefore, if the benefit we find for unconventional referents is due to knowledge of location, then we should find the same effect for conventional referents. Any additional benefit for unconventional objects, over and above the effect for conventional objects would constitute a true benefit due to linguistic precedents. Our first experiment seeks to confirm the existence of this differential benefit.

Method

Participants. Twenty college students (12 males and 8 females) from the University of Chicago, who were all native speakers of American English, participated as addressees in the experiment for payment. Nine additional participants did not provide useable data due to calibration problems or problems with the recording equipment.

Apparatus. The director and addressee sat at a small table, facing one another. Between them we placed a vertical grid made of opaque white fiberglass. The grid was composed of a set of 16 boxes arranged in a 4 × 4 pattern (see Fig. 1). The measurements for each square in the grid were 12.5 × 12.5 × 12.5 cm. The squares were uncovered, so it was possible to see through each square to the other side. Thus, objects in the grid were mutually visible to both people.
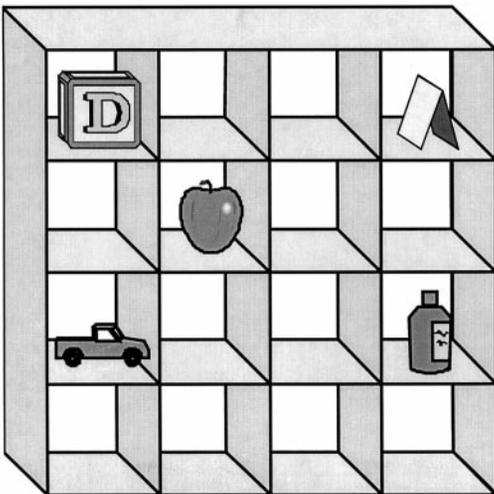

FIG. 1. Example of an Experimental Item in Experiment 1.

We used an Applied Science Laboratories (ASL) Series 4000 head-mounted eyetracking system to monitor the position of the addressee's left eye. The addressee wore a headband on which an eye camera and a magnetic head tracker were mounted. This setup permitted addressees to move their heads about freely while performing the task. The eye camera provided data concerning the position of the participant's eye relative to the head, while the head tracker corrected for movements of the head. Together, these two values determined the position of the participant's gaze.

A free-standing video camera filmed the grid from the addressee's point of view. The eyetracker superimposed a crosshair on the video image corresponding to the position of the addressee's gaze at a temporal resolution of 30 Hz, or approximately 1 sample every 33 ms. Additionally, the real-valued coordinates of the addressee's gaze relative to the grid were stored digitally by a PC running ASL E4000 software at a rate of 60 Hz. Two microphones placed on the table recorded the conversation onto the videotape.

Procedure. To ensure that all addressees heard the same critical utterances during the experiment, the director was a trained (male) confederate who produced scripted utterances in a seemingly spontaneous manner. To enhance the realism of the experimental situation, we used several credibility cues throughout the experiment. The confederate arrived 5 min late to the experiment and pretended never to have met the experimenter. Later, during a set of practice trials (described below), the confederate made a few errors that a naïve subject might have made.

Once the confederate arrived, the experimenter told the pair that the experiment would involve two different roles, that of director and addressee. We then staged a "random" assignment of roles. The experimenter asked the pair to choose from among two slips of paper that were facing down that were supposedly marked with the two different roles. In reality, both were marked "Addressee." In this way, the real participant always drew the role of addressee, and the confederate simply announced that he had received the role of director.

The experimenter then told the pair that they would be playing a series of communication games where the goal was to rearrange objects in a grid to match a goal state. The director would be given a photo showing what needed to be done to reach this state. The addressee would not be allowed to view the photo and the director could not move the objects. To reach the goal state, they had to work together, with the director instructing the addressee to move objects and the addressee following the director's instructions.

We designed the next portion of the instructions to lead the participant to believe that the director would be producing the instructions spontaneously. The experimenter explained that at the beginning of each trial, he would give the director a card containing a photograph of the grid. The cover story was that the card contained a photograph of the grid in its initial state, with arrows indicating how objects were to be moved. Therefore, the director would only know which objects needed to be moved and where they should go, but would have to come up with his own way of describing the objects. Unbeknownst to the addressee, the director actually had scripted instructions printed on the cards.

We pretrained the confederate to avoid using eye gaze to communicate the position of the target. The confederate looked back and forth between the card and the grid while preparing each utterance, making sure that the target was not the last object that he looked at. While delivering the critical utterance, he kept his eyes fixated on the card. After the utterance, he was allowed to look up at the addressee.

We used three practice trials to make sure that the addressee understood the task and believed the cover story. The director's cards for these trials contained actual photographs of the grid with arrows showing how objects were to be moved. At the end of the first two trials, the experimenter showed the addressee the director's card with the arrows so that he or she could see that there were no written labels for the objects. After completing two practice trials, the partners switched roles. Playing the role of the director made it extremely clear to the participant/addressee that the director would have to come up with labels on his own.

The experimenter then introduced the eye-tracking equipment and explained its operation. To avoid drawing undue attention to eye movements, the pair was told that the eye camera recorded features of the eye "such as pupil diameter." Finally, the experimenter mounted the eyetracking equipment on the addressee's head and performed the calibration procedure.

Next, the experimenter gave the participants two final instructions. First, he instructed them not to talk about or touch the objects, except as the task required. This was mainly to prevent the addressee from spontaneously naming objects (e.g., "look at this cute toy monkey!") before the director could refer to them. Second, he told the confederate to say "Ready" before each instruction. This was a cue for addressees to look at the center position on the grid, so that for every trial their gaze would start in the center.

After completing all of the experimental trials, the experimenter gave both participants a written questionnaire with general questions about the experiment. It ended with a question that probed whether the addressee suspected that the director was a confederate. The question stated that for some of the pairs in the experiment we used a trained confederate and offered financial incentive for the participants to accurately guess whether their partners were confederates. After completing the experiment, the experimenter fully debriefed participants about the experiment and allowed the participants to ask any questions. Of the 20 participants, only 3 (15%) guessed that the director was a confederate.

Materials and design. We performed an informal norming study to select objects with and without conventional names. We presented 22 native English speakers with 50 objects, one object at a time. Each participant was asked to name the objects as they would if they had to command someone to "Pick up the _____." We drew 12 conventional referents from the set of objects that generated agreement on names for at least 70% of participants, and 12 unconventional referents from those that generated agreement for less than 30%. These referents were used as target objects in the experiment. For conventional targets, the names that we chose were those that respondents had used most frequently. For unconventional objects, the respondents produced highly idiosyncratic descriptions. Therefore, instead of choosing the most frequent name, we chose names that we thought were not easily predictable but specific enough to uniquely identify the object from among the set of alternatives in the grid. Names and a short description for all 24 targets are listed in Table 1.

TABLE 1

Conventional and Unconventional Referents, Experiment 1

Conventional objects		Unconventional objects	
Basket:	A small wicker basket.	Link:	A metal link open at one end.
Crayon:	A Crayola crayon.	Belt:	A pink velcro strap or belt.
Straw:	A drinking straw.	Foam:	An odd-shaped piece of green foam.
Eraser:	A white pencil eraser.	Probes:	A pair of multimeter probes, with wires coiled and rubber banded.
Candle:	A small green candle.	Adapter:	A "Y" adapter which allows two sets of headphones to be plugged into one jack.
Pen:	An ordinary ball-point pen.	Wires:	A hanger for collectible plates, composed of two springs and two bent pieces of wire.
Pencil:	An everyday pencil.	Hook:	A metal coat hook typically found mounted on the back of doors.
Soap:	A bar of soap, still in its wrapper.	Razor:	A small retractable knife.
Chapstick:	A tube of Chapstick.	Mustard:	A small packet of mustard.
Sunglasses:	A pair of men's sunglasses.	Wrench:	A piece of a door latch that looks "wrench-like".
Matches:	A bundle of wooden matches.	Nail:	A small eye-screw.
Battery:	An AA battery.	Tent:	A piece of paper tri-folded to make an inverted "V" shape.

We embedded these target objects among filler objects in the grid to create items for the experiment. Each experimental item consisted of a preparatory instruction and a critical instruction. The preparatory instruction always preceded the critical instruction, and either grounded the target object by mentioning it as a landmark or left it ungrounded. In Fig. 1, the target object is the folded paper ("tent") in the top right-hand corner. For this grid, the preparatory instruction in the grounded condition might be "the toy truck goes below the tent." To attain maximal consistency across conditions, when the preparatory instruction left the target object ungrounded, it required the addressee to perform the same action as in the grounded condition, though without mentioning the target object. For Fig. 1, in the ungrounded condition this instruction would be "the toy truck goes above the bottle." Thus, in both conditions, the instruction requires the addressee to move a filler object (e.g., the toy truck) to the same square, though the target object is grounded in only one condition.

After the preparatory instruction came the critical instruction, occasionally with a filler instruction intervening. The critical instruction for each item was always the same. For the grid in Fig. 1, the critical utterance was, "the tent goes below the apple." We compared how long it took addressees to identify the tent in the grounded and ungrounded conditions.

Between trials, the experimenter would replace objects from the previous trial with objects for the next trial. To minimize setup time between trials, each grid contained two experimental items. The two items within a single grid always appeared in opposite conditions. For instance, if one item was unconventional and grounded, the other item would be conventional and ungrounded. The preparatory and critical instructions for the two items were interleaved in each grid, and some grids had additional "filler" instructions to obscure the purpose of the experiment.

The design of the experiment was a 2 (Conventionality: Conventional vs Unconventional) × 2 (Mention: Precedent vs No Precedent) within-participant design.

Coding and analysis. An undergraduate research assistant, who was blind to the hypothe-

ses of the study, coded the videotapes. For each critical utterance, the coder located two points that defined response time. The first point was the onset of the target word, defined as the initial syllable of the name of the target object (e.g., the first syllable of "tent"). The last point was the "decision point" or "final fixation," defined as the point at which the addressee identified the target object as the referent. This was operationalized as the beginning of the addressee's last fixation[1] on the target object before touching it. We also extracted the latency of addressees' first fixation on the target from the digital data. In some cases, upon locating the referent addressees would immediately identify and move the target. In these cases, the same fixation was both first and final.

Additionally, the coder noted whether the addressee asked the director for clarification or confirmation about the referent. For instance, the addressee might ask, "is this the tent?" This was noted because such overt exchanges would possibly inflate the response time in the ungrounded condition and thereby overcontribute to the effect.

Results and Discussion

Response times for four trials (0.8% of the data) could not be computed because the addressees used peripheral vision and did not look at the target square or the crosshair failed to enter the square due to poor calibration. We removed these data points from the analysis. Additionally, to reduce any inflation in response time due to clarification or confirmation requests, we truncated response times, using the top of the distribution of items where no such exchanges occurred as the cutoff point. This truncation procedure affected only 2.3% of the data. Complete means and standard deviations for the measures are provided in Table 2.

In general, addressees took longer to identify referents when no precedent existed (the means were 3018 ms in the No Precedent condition and 1710 ms in the Precedent condition). In other words, the existence of a precedent shortened

---

[1] The criterion for a "fixation" was that the eye must remain within a particular square of the grid for 100 consecutive ms, or three video frames.

TABLE 2

Mean Fixation Latencies (and Standard Deviations) by Mention and Conventionality, Experiment 1

	First fixation		Final fixation	
	No Precedent	Precedent	No Precedent	Precedent
Conventional	1453 (940)	1331 (800)	2101 (1655)	1725 (1074)
Unconventional	1701 (1117)	1246 (837)	3950 (2921)	1695 (1093)

referential search by an average of 1308 ms. We submitted the data to an analysis of variance and report subject and item analyses as $F1$ and $F2$, respectively. As predicted, we found a main effect of mention, $F1(1,19) = 45.10$, $p < .001$, $MSE = 779123$; $F2(1,11) = 55.17$, $MSE = 368153$, $p < .001$.

More importantly, this benefit due to an existing precedent was much more pronounced for unconventional (2255 ms difference) than for conventional objects (376 ms difference). The interaction was significant, $F1(1,19) = 59.30$, $p < .001$, $MSE = 303140$; $F2(1,22) = 17.97$, $MSE = 587876$, $p < .01$, as were the individual precedent effects for unconventional referents, $t(19) = 7.88$, $p < .001$, and conventional referents, $t(19) = 2.39$, $p < .05$.

Analysis of the first fixations reveals the same basic pattern. We lost the digital data for three participants, so they were excluded from this analysis. Addressees took longer to first fixate on referents when there was no precedent (1577 versus 1289 ms), yielding a benefit of 288 ms. The main effect of mention was significant, $F1(1,16) = 5.26$, $p < .05$; $MSE = 852$; $F2(1,22) = 14.42$, $MSE = 304$, $p < .01$. The interaction, however, was only marginal but in the predicted direction, $F1(1,16) = 2.70$, $p < .13$, $MSE = 413$; $F2(1,22) = 2.49$, $MSE = 758$, $p < .13$.

The finding that precedents provide a larger benefit for unconventional referents excludes the possibility that the effect was due to knowledge of location of the referent. The larger benefit for unconventional referents was mainly due to a long search time when they were unmentioned (3950 ms on average). Unconventional and conventional referents that had been mentioned were statistically indistinguishable from one another [mean RTs 1695 and 1725 ms, respectively, $t(19) = 0.29$, ns]. In effect, the linguistic precedent made the unconventional referents look conventional.

In sum, linguistic precedents produce greater benefits for comprehension when a referent is unconventional than when it is conventional. What mechanism is responsible for this benefit? One possibility, consistent with the standard theory, is that linguistic precedents benefit comprehension because they are part of a mutually known background of information. Another possibility is that they benefit comprehension independently of mutual knowledge, simply because they are available to the addressee.

The collaborative model of Clark and colleagues (Clark & Wilkes-Gibbs, 1986; Clark & Brennan, 1991) implies that listeners should exhibit more certainty when they interpret a label for which a mutually known precedent exists than when the precedent exists but is not mutually known. This implication stems from two sources.

First, it is claimed that speakers and listeners collaboratively "ground" the meanings of linguistic expressions. One of the consequences of grounding is the creation of the mutual belief that some term or utterance has been properly understood (Clark & Wilkes-Gibbs, 1986; Clark & Brennan, 1991). It follows that listeners are only truly justified in applying a linguistic precedent to understand subsequent uses of the term when this mutual belief exists. When this belief is lacking, such as when they interpret the utterances of a new speaker who was not co-present when the original precedent was established, they should seek to once again ground the term because the meaning of the term would be less certain.

The second way in which the collaborative model implies that grounding is most effective

in relation to mutual knowledge originates in claims concerning the nature of the comprehension system itself. Clark and Carlson (1981) suggest that when addressees comprehend utterances, they should restrict the information they consider to mutual knowledge, or common ground. It follows that the comprehension system should ignore any precedents that exist outside of this shared body of knowledge. This would prevent it from making otherwise systematic errors, such as when Henry interpreted "I've been reading the reports" as referring to the reports about his wife.

Unlike the common ground model, Perspective Adjustment predicts that listeners use precedents simply because prior use has made them egocentrically available. If this is the case, then listeners should be no less likely to use a precedent when it is not part of common ground than when it is. Experiments 2 and 3 use multiple speakers to test these two models.

EXPERIMENT 2

In this experiment, we used the same task as in Experiment 1, except that addressees listened to instructions from not one, but two directors: a live or "present" confederate director as in Experiment 1 and a "recorded" director, whose instructions the addressee heard through a headphone. The fact that the present director could not hear the recorded director's instructions ensured that the addressee would have to establish separate precedents with each director. In the experiment, a description for a referent was mentioned first by one of these two directors and then referred to in a target instruction by the present director. In other words, while there was always a precedent for the second mention, that precedent had been established by either the present or recorded director. The question was: Would comprehension benefit more from a precedent when it was used by the same director who had established it than when it was used by a director who had not? Alternatively, would the benefit be independent of the identity of the director who established the precedent?

We designed the experiment to distinguish among three specific hypotheses: (1) *strong partner specificity*, (2) *weak partner specificity*, and (3) *partner independence*. The strong version of the partner-specificity hypothesis predicts that the existence of a precedent will *only* produce a benefit when the director that refers to a test object is the same one who originally mentioned it. Because in our experiment the present director always refers to the test object the second time, the strong partner-specificity hypothesis predicts a benefit only when the present director was the one who originally established the precedent. A weak version of this hypothesis predicts that the identity of the director will matter somewhat. The idea that mutual knowledge can be a probabilistic constraint on comprehension was advanced by Hanna, Trueswell, Tanenhaus, and Novick (1997). It predicts the largest benefit when a precedent exists and is part of mutual knowledge; that is, when the present director is the one who uses the label first. There should still be a benefit when a precedent exists, but is not part of mutual knowledge; that is, when the referent has been mentioned previously, but not by the present speaker. However, this benefit should be smaller than in the case where the target was mentioned both times by the same speaker. In contrast to these two hypotheses, the partner-independence hypothesis predicts the exact same benefit irrespective of the identity of speaker who established the precedent.

Method

Participants. Thirty-six college students from the University of Chicago, who were all native speakers of American English, participated as addressees in exchange for payment. Seventeen additional participants did not provide useable data due to poor calibration, experimenter error, or problems with the audio equipment.

Apparatus. We used the same eyetracking setup as described in the previous experiment. In addition, addressees wore a headphone covering one ear, where they heard spoken instructions from a "secret" prerecorded director. The instructions were recorded by a male undergraduate student who was a native English speaker. These instructions had been stored as digital audio files on the hard drive of a PC compatible computer equipped with a sound card. The pres-

ent, confederate director activated the sound files using a computer keyboard.

Procedure. In many ways, the procedure for this experiment was similar to that used in Experiment 1. Therefore, in this section we focus mainly on changes to the procedure that were introduced to accommodate the use of the second (recorded) speaker.

As in the previous experiment the director was a confederate, a female undergraduate student from the University of Chicago who was a native speaker of American English. We chose a male speaker as the recorded director to contrast with the female confederate. To motivate the use of both a present director and a prerecorded director we informed participants that we were researching how people respond to live versus recorded speech.

As in the previous experiment, we used the cover story that the director would receive cards containing photographs of objects, with arrows indicating how they were to be moved. The experimenter explained that the arrows on the picture were marked to indicate who would deliver each instruction, the present or recorded director. Some of the arrows would be marked with letters (e.g., A, B, AA, and AB) to indicate that the corresponding instruction was a secret instruction to be delivered by the recorded director. The director entered the letters on a computer keyboard to activate the playback of a sound file that could be heard only through the addressee's headphone. When an arrow was not labeled, this meant that the instruction was to be given by the present (confederate) director herself. In actuality, the director would be reading scripted utterances printed on the cards and activating recorded instructions when necessary. As in Experiment 1, the confederate was trained not to indicate the target using her eyes.

As in the first experiment, we used three practice trials, including a third role reversal which made it clear to the addressee that (1) the director had a picture with arrows and needed to generate labels for objects and (2) the director would be unable to hear the instructions from the recorded director because the addressee would hear them through an earphone.

After the practice trials, the eyetracking

equipment was mounted, calibrated, and the experiment began. Before the first experimental grid, the partners completed a warm-up grid.

As in the previous experiment, a postexperiment questionnaire examined whether participants thought that the director was a confederate.

Materials and design. The target objects were the 12 unconventional referents from Experiment 1, except that we replaced one of the critical objects because it turned out to be too small and therefore somewhat difficult for addressees to locate.[2] We created 12 experimental items by embedding these referents in a grid with filler objects. As in the previous experiment, each experimental item consisted of one preparatory instruction and one critical instruction. Each grid contained two experimental items, which always appeared in opposite conditions. For instance, if one item was mentioned by the present director, the other item was unmentioned and the preparatory instruction for that item was given by the recorded director.

The experiment had two, two-level within-participant factors: Mention (Precedent vs No Precedent) and Precedent Speaker, the identity of the speaker who set the precedent (Present Director or Recorded Director). We created four different versions of the experiment, so that each item appeared in all of four conditions.[3] The order of the grids was randomized for each participant.

Coding and analysis. As in Experiment 1, a coder who was blind to the hypotheses of the study located the beginning of the critical word on the videotape for each experimental trial. We also used the same criteria to locate the final fixation on the target object and extract the first fixation from the digital data.

---

[2] Specifically, we replaced the "nail" (see Table 1) with a large blue soap shaped like Batman.

[3] Unfortunately, the distribution of conditions within two versions of the items was not perfectly balanced. This was due to a programming error in the software that generated the four versions of the items. To correct for possible artifacts, in addition to our overall analysis we ran statistical analyses on the two versions of the items that were properly balanced. All effects that were significant in the overall analysis were also significant in the analysis that excluded the two unbalanced versions.

Results and Discussion

Data from 25 trials (5.48%) could not be used because there was no final fixation on the target due either to peripheral vision or poor calibration. We excluded 12 trials from the analysis where addressees moved the wrong object (2.63% of the data) and three trials where there was experimenter error in setting up the grid (less than 1%). In addition, there were a few very long search times in this experiment (longer than 10 s), so we truncated response times at the 98th percentile (approximately 10 s).

As in Experiment 1, unmentioned referents took much longer to identify than mentioned referents (the means are 4374 and 2902 ms, respectively).[4] This main effect of mention was significant, $F1(1,35) = 83.72$, $MSE = 910645$, $p < .01$; $F2(1,11) = 16.45$, $MSE = 1532168$,

[4]It may be noted that the response times in Experiment 2 appear to be slightly longer than those observed in Experiment 1. This is mainly attributable to the greater number of objects in the grids in Experiment 2, where each grid had an average of 15 objects, versus 11 in Experiment 1. Therefore, addressees in Experiment 2 had more objects to search among, resulting in longer search times.

$p < .01$. There was no main effect of the identity of the director nor interaction with mention (all Fs ≈ 1).

Figure 2 presents the results for the final fixation on the target (means, No Precedent vs Precedent: Present Director, 4428 vs 2984; Recorded Director, 4319 vs 2819). It clearly shows that the benefit of a linguistic precedent was the same regardless of who had originally mentioned the referent, with a benefit of 1444 ms when the referent was mentioned by the same (present) director and of 1500 ms for the different (recorded) director. There was no hint of an interaction between mention and the identity of the first speaker ($F1$ and $F2 < 1$); indeed, the 56 ms difference in benefit was in the opposite direction of what would be predicted by either of the partner-specificity hypotheses. When addressees could not assume mutual knowledge of the precedent, they were no less likely to use it than when they could assume mutual knowledge. These results unambiguously support the partner-independence hypothesis: The benefit of the precedent is due to its availability rather than the fact that it is mutually known.

The first fixation measure also failed to reveal any effect of the identity of the first speaker on
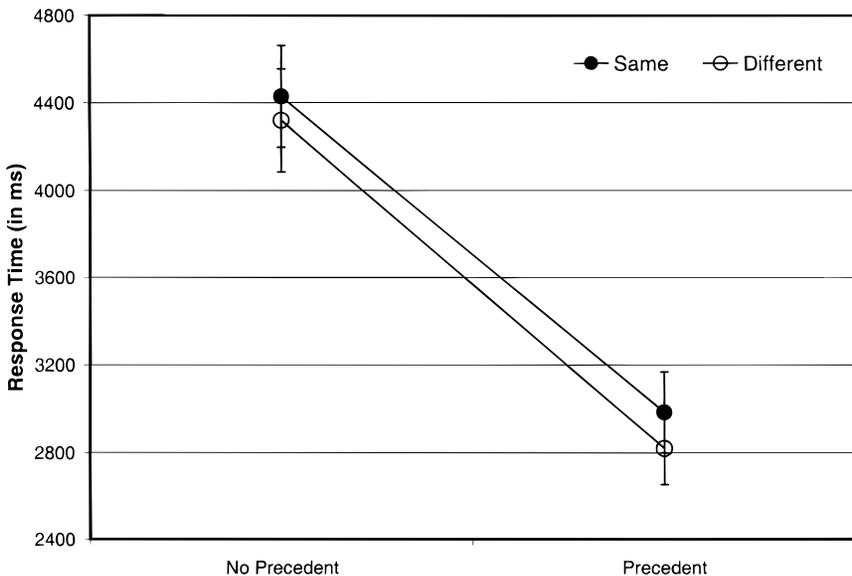

FIG. 2. Fixation Latency × Precedent and Precedent Speaker in Experiment 2.

initially locating the target object. Latencies for the first fixation were truncated at the 98th percentile (5833 ms). When the precedent speaker was the present director, the mean was 1529 (vs 2098 when there was no precedent). When it was the recorded director, the mean was 1449 (vs 2013, no precedent). While the difference in benefit was only 5 ms, the difference of 80 ms (1529 vs 1449) between the two Precedent conditions was in the opposite direction of what partner specificity would predict. Therefore, mutual knowledge of linguistic precedents does not seem to impact the early moments of comprehension.

It is possible that we found no evidence for partner specificity because some participants realized that the present director was a confederate. Perhaps these participants treated both directors as if they were drawing upon the same mutual knowledge. The postexperiment questionnaire revealed that 47% of participants in the experiment guessed that the director was a confederate. Most of these participants (65%), though, reported that this possibility did not cross their mind during the experiment. To evaluate whether guessing the identity of the confederate made a difference, we compared the data of participants who guessed with those of participants who did not guess. If partner specificity depends on participants' belief that the present director is a true participant and not a confederate, then one would expect to find more evidence for partner specificity with participants who did not guess than with participants who did guess that the director was a confederate.

However, no such pattern emerged. In fact, we saw an opposite trend: The benefit for nonguessers was 150 ms larger on average when the recorded director first mentioned the term than when the present director did. The factor of whether addressees correctly guessed failed to explain any variation in the differential benefit ($r^2 = 0.004$, *ns*). These findings demonstrate a solid pattern of partner independence.

The fact that we found no effect of speaker identity on the length of time needed to identify a mentioned referent is surprising, in that it indicates that the comprehension system ignores a potentially valuable source of information.

Though the meanings of referential expressions are less certain when a mutually accepted precedent does not exist, addressees seemed indifferent to the mutuality of the precedent. From a theoretical standpoint, it is possible that reasoning about common ground is a largely deliberative process that is too slow to constrain the rapid and relatively nonreflective processes of referential search (Keysar et al., 2000).

A possible objection is that it was too difficult for addressees to keep track of the two perspectives because of the intermixing of utterances from two different speakers, a circumstance that some might regard as unnatural. However, a natural analogy to the experimental situation would be when someone talks to one person over the phone and another person, who is present in the room, occasionally interjects a side conversation. We think that it is implausible that under such circumstances an addressee could not, in principle, easily distinguish between the perspectives of their two conversational partners. Still, we went to great lengths to clearly accentuate the differences between the speakers. First, we required addressees to play the role of the present director in a role-reversal practice trial. This permitted them to observe that the present director would not hear the recorded descriptions. Second, the voices of the two directors were very different because of the difference in gender. Third, at the beginning of the experiment it was emphasized to participants that the recordings were *secret* instructions that they should not reveal to the present director. That we found no evidence for the mutuality of precedents despite these procedures supports the idea that these processes operate independently of common ground.

A similar objection is that perhaps addressees *inferred* mutual knowledge between the two speakers to make sense of what may have seemed an improbable event: two different speakers repeatedly using the same unconventional names to describe targets. If this is the case, then listeners would have stronger evidence for mutual knowledge between directors in later trials than in earlier trials. According to this explanation, we should find a partner-specific effect only in the beginning of the experi-

ment, before the addressee had evidence that the directors would use the same descriptions. To test this, we considered only the first trial for each subject in each of the four conditions. The means for No Precedent vs Precedent were 4696 vs 2876 in the Recorded Director condition, or a benefit of 1820 ms; in the Present Director condition, the means were 4696 vs 2978, or a benefit of 1832 ms. While there was a significant main effect of precedent, $F(1,35) = 51.63$, $MSE = 2652393$, $p < .001$, there was no trace of a partner-specific effect ($F < 1$).

A final concern might be that addressees had to remember the locations of referents from among a large array, leading to some long search times that may have masked any early effect of mutual knowledge. To address this concern, along with the others mentioned above, in Experiment 3 we replicate our findings using a simpler task.

Experiment 3 explores the additional issue of overspecification. The more frequently speakers refer to a referent, the more they tend to standardize their description (Clark & Wilkes-Gibbs, 1986; Krauss & Weinheimer, 1964, 1966). This can cause them to overspecify referents, such as when they call the only shoe in a display of objects a "loafer" (Brennan & Clark, 1996). When speakers overspecify referents, they violate Grice's Maxim of Quantity (Grice, 1975) because they convey more information than necessary. This raises the interesting question of how listeners respond to this overspecification. If listeners expect speakers to use precedents, then comprehension should be facilitated. If they expect speakers to adhere strictly to the Maxim of Quantity, then overspecification vis-à-vis precedents should impair comprehension.

EXPERIMENT 3

The motivation for this experiment was twofold: (1) to investigate whether listeners expect speakers to overspecify referents due to precedent use and (2) to test the issue of partner specificity, i.e., whether listeners are more likely to expect a speaker to use a precedent when they have a basis for inferring that the precedent is shared. As in the previous experiments, this ex-

periment explored these questions using eye-tracking.[5]

In addition, this experiment improved on the design of Experiment 2, offering greater sensitivity to small, early effects of mutual knowledge and a more compelling separation of speakers. There were only two images for listeners to select from, and listeners could view them simultaneously, reducing memory demands. Utterances from the two speakers were not interleaved but blocked, following Brennan and Clark (1996), who investigated speakers' use of precedents by blocking trials with different addressees. Thus, the identity of the speaker changed at most once during the entire experiment, unlike Experiment 2, where listeners had to switch back and forth between two speakers with whom they had different common ground. Furthermore, speakers used preexisting conventional names to identify referents, avoiding the potential problem in Experiment 2 where two different speakers used the same improbable names.

In the experiment, pairs of pictures of conventional, everyday objects appeared on a computer screen. Listeners heard a speaker name of one of the pictures and selected the corresponding picture by clicking on it with the computer mouse.

During the experiment, listeners heard a female speaker (Speaker A) establish subordinate-level precedents for pictures of everyday objects, as speakers did in Brennan and Clark (1996). For instance, listeners heard her call a picture of a car a "sportscar" when it appeared in the context of a station wagon, and a picture of a flower a "carnation" in the context of a daisy. Speaker A "entrained" on the subordinate-level precedents by using them multiple times in these contexts over the course of the experiment.

After establishing precedents, at the end of the experiment listeners completed a set of

---

[5] We also tracked movements of the computer mouse and found results that were closely time-locked to the eye-movement data. Mouse tracking presents an easy-to-use, low-budget alternative to eyetracking. Information about the mouse tracking data and the technique of mouse tracking can be obtained by contacting the first author.

"posttest" trials where target pictures appeared in contexts where basic-level names would be appropriate. These trials probed whether listeners' experience with the precedents led them to expect to hear subordinate-level overspecification or conventional, basic-level names. For instance, the carnation appeared in the context of the sportscar. Thus, listeners could expect either the basic-level names "flower" and "car" or the subordinate-level precedents "carnation" and "sportscar."

In these trials, the speaker used the basic-level name for the target referent, e.g., "car." Note that "car" overlaps phonologically with the onset of the subordinate-level precedent for the other picture, "carnation." Therefore, if listeners are expecting the subordinate-level precedents "sportscar" and "carnation" instead of "car" and "flower," then when they hear "car," the "carnation" precedent should interfere with selection of the car as the referent.

This experiment tested the strong and weak partner-specificity hypotheses against partner independence. All listeners first entrained on the precedents with Speaker A. Later, half of the listeners completed the posttest trials with Speaker A, while the other half completed them with a new, male speaker, Speaker B. Listeners who continued with Speaker A (Same Speaker condition) could infer mutual knowledge of the precedents through linguistic and social copresence (Clark & Marshall, 1981). However, listeners who switched to Speaker B had no evidence for their copresence and could not infer mutual knowledge.

In order to ascertain the interference due to precedents, the posttest performance of both groups was compared to their performance on these same items during an earlier "pretest" with Speaker A. The pretest was conducted at the beginning of the experiment, before the subordinate-level precedents were established, and therefore provides an important baseline. As in the posttest, listeners saw the car and flower on the screen and heard "car." Strong partner specificity, where listeners fully constrain comprehension to mutual knowledge, predicts that listeners who hear Speaker B say "car" in the posttest should be no more likely to look at the carnation

than they were when they heard "car" in the pretest. Likewise, if listeners who continue with Speaker A use precedents, then when they hear Speaker A say "car," they should experience interference from the precedent "carnation." That is, they should exhibit a higher probability of looking at the carnation than in the pretest. The weak version of partner-specificity predicts that listeners in the Different Speaker condition will still show interference, but less than those in the Same Speaker condition. The partner-independence hypothesis predicts equally strong interference in both conditions.

Unlike Experiment 2, the spoken utterances used as stimuli in this experiment were all prerecorded. This is because real speakers would be somewhat unlikely to produce the precise descriptions with overlapping onsets needed to test the hypotheses. Moreover, the use of prerecorded materials permitted stringent matching of the timing and duration of the two speakers' posttest utterances to reduce variability and increase the likelihood of detecting small effects.

We were concerned about the possibility that prerecorded materials would compromise the generality of our findings. This is because theories of language use make a fundamental distinction between addressees and overhearers (Clark & Carlson, 1982; Schober & Clark, 1989). Listeners who know that they are listening to prerecorded materials would believe themselves to be "overhearers" of utterances that were not designed for them. For this reason, we strove to convince one set of participants that they were actually listening to a real speaker who spoke to them, live, from another room. In other words, these listeners believed themselves to be nonparticipating addressees because they could not interact with the speaker (Addressee condition). From a research standpoint, this introduced undesirable aspects into the procedure because it requires deception and elaborate techniques to incorporate credibility cues. In the interest of simplifying future research, we compared the Addressee condition to the condition where we told listeners that they would simply overhear prerecorded utterances from real participants (Overhearer condition). If we find no difference, this suggests that future research on

precedent use may not require such elaborate procedures.

Finally, as in Experiment 2, we used a postexperiment manipulation check to examine the success of the cover story. We also added a question and a recall test to the Different Speaker condition to investigate whether listeners treated the perspectives of the two speakers differently.

Method

Design. The experiment used a mixed design consisting of two two-level between-participants factors, Speaker (Same or Different speaker in the posttest) and Listener Status (Addressee or Overhearer). There was one two-level within-participant factor, Test (Pretest or Posttest).

Participants. Sixty-four undergraduates from the University of Illinois at Urbana–Champaign participated for course credit. There were 32 males and 32 females, equally distributed among the four cells that result from combining the factors of Speaker and Listener Status.

Apparatus. A computer controlled the presentation of stimuli and playback of the prerecorded instructions, synchronizing these events with the recording of eyetracking data. We used an Eyelink eyetracking system instead of the ASL eyetracker used in the previous experiments. The main difference between the two systems is that the Eyelink system acquires data at a sample rate of 250 instead of 60 Hz.

General procedure. In this section, we first provide a general overview of the experiment for all conditions, except where noted. Then, we provide details about the two Listener Status conditions.

Participants arrived at the laboratory and received instructions, and the experimenter introduced the cover story (details below). The participant sat at a comfortable distance in front of a computer screen and operated a computer mouse. After operation of the eyetracker was explained to the participant, the experimenter put it on the participant's head and completed a brief calibration procedure.

In each trial, two images appeared on a computer screen, spaced at an equal distance from the center of the screen. Listeners were permitted to freely view the images for 1 s. We decided to use this free-viewing technique instead of forcing participants to stare at a central fixation point because a pilot study revealed that the fixation point technique encouraged participants to select targets using only their peripheral vision. After the 1-s inspection period, listeners heard the speaker name one of the two images, and they clicked on the referent with the mouse. They did not receive feedback on their performance. After a picture was selected, the screen was cleared and the next trial began.

For readers' convenience, a flowchart of the key aspects of the experiment is presented in Fig. 3. The temporal order in which the different items appeared follows the flowchart from left to right. For simplicity, the figure omits three practice trials at the beginning, some filler trials, and the recall test and questionnaire at the end of the experiment.

It is important for the reader to keep in mind that our division of the experiment into different phases ("pretest," "entrainment," "posttest," etc.) is for expository purposes only and does not correspond to different phases of the listeners' experience. Listeners were not aware of these divisions, experiencing only a sequence of trials that was interrupted once, just prior to the pretest, to accommodate the introduction of Speaker B in the Different Speaker condition.

The experiment began with three practice trials (not shown in Fig. 3). The images in the practice trials were not seen again during the experiment. Then came eight pretest items, which provided a comprehension baseline before precedents were established. The items were presented in a random order. Listeners heard the speaker use basic-level terms to name targets (e.g., "car" and "butterfly"). Because listeners would see these displays again, we wanted to make sure that the two pictures of each pair were referred to with roughly equal frequency, so that there were no "favored" targets. For this reason, immediately following the pretest came a set of filler items (not shown in Fig. 3) where the same displays appeared, but the listener heard the eight opposite pictures referred to, also using basic-level terms (e.g., "flower" and "knife").
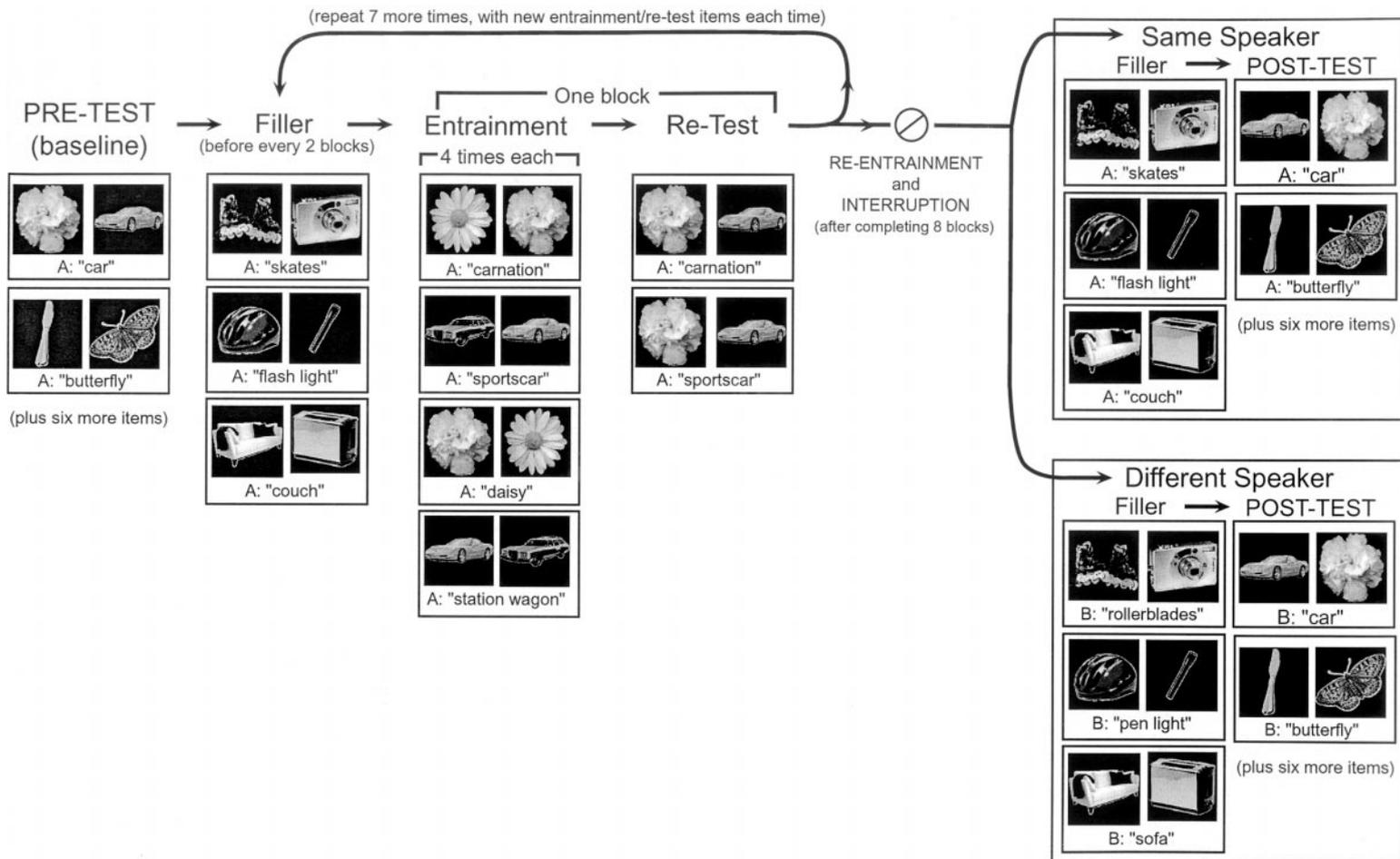
FIG. 3. Design of Experiment 3. "A" and "B" refer to Speaker A and Speaker B. The figure is schematic and excludes practice trials, certain filler items, the reentrainment phase with Speaker A prior to the posttest, and the postexperiment recall test.

After the pretest, listeners completed eight blocks of items where they entrained on subordinate-level precedents. Every block presented different pictures and different subordinate-level terms to entrain upon. The order of the blocks was randomized. Before every two blocks came the same three filler items (shown in Fig. 3), that we later used in the recall test.

In each block, the two images from each pretest item appeared as targets four times in contexts that now required subordinate- instead of basic-level names. See the figure for an example of four entrainment displays. For instance, the flower appeared with another flower and was called "carnation"; the car, which appeared with another car, was called "sportscar." Listeners also heard the speaker refer to the new flower ("daisy") and new car ("station wagon") four times each. The 16 trials for each block of the entrainment phase were presented in a random order.

Each block ended with a retest item to probe whether listeners would expect to hear precedents in a basic-level context. Figure 3 shows the retest item for one block, in which the listener sees the car and flower and hears "carnation." For this item, the question was whether listeners would expect "flower"/"car" or "carnation"/"sportscar." Right after the probe item, listeners saw a similar display and heard the other picture, the car referred to (as "sportscar"). Again, this was so that both pictures would be referred to equally often. Listeners' performance on retest items is not crucial to our hypotheses and is redundant with the posttest items; therefore, in our results we focus mainly on posttest items.

In the next block, listeners would entrain on precedents for a different set of pictures (e.g., "monarch," "butter knife," "black butterfly," and "steak knife") and then complete two retest trials for that set. Because each block introduced four new precedents, for a total of 32, there was a concern that listeners might have forgotten the precedents from early blocks by the time of the posttest. To remind them, after the last block of entrainment and just prior to the interruption, participants underwent "reentrainment" on the precedents. The eight images that would be used

as subordinate-level competitors (e.g., "carnation" and "butter knife") in the posttest phase appeared one time each as targets with their same category images (e.g., "daisy" and "steak knife"). These appeared in a random order.

To summarize, listeners completed 183 trials with Speaker A. There were 3 practice trials; 8 trials in the pretest, plus 8 fillers immediately following; 12 filler trials (3 before every two blocks); 144 entrainment trials (16 entrainment and 2 retest for each block); and 8 reentrainment trials. Listeners required approximately 20 min to complete Speaker A trials.

Following completion of Speaker A trials, there was a 3-min "interruption" in both the Same and Different Speaker conditions. This was to accommodate the introduction of Speaker B in the Different Speaker condition. The details of this procedure differ for the two Speaker and two Listener Status conditions and are provided in the respective sections below.

The posttest was preceded by three filler trials, the same displays that had appeared before every two blocks with Speaker A. Half of the listeners completed these fillers and the posttest items in the Different Speaker condition. In this condition, we used the filler trials to emphasize to listeners that Speaker B did not know Speaker A's precedents before listeners completed any posttest items. To this end, Speaker B named the filler targets in a way that was inconsistent with Speaker A's precedents. Whereas Speaker A had called these targets "skates," "flashlight," and "couch," Speaker B called them "rollerblades," "penlight," and "sofa." In the Same Speaker condition, Speaker A continued using her precedents for the fillers.

Next, listeners in both conditions went on to complete the eight posttest trials, which appeared in a random order. Listeners in the Different Speaker condition completed them with Speaker B, and listeners in the Same Speaker condition, with Speaker A. Note that by the time of the posttest, listeners had heard the basic-level terms "car" and "flower" one time each (in the pretest), the precedent "sportscar" five times (four during entrainment and once in the retest) and "carnation" six times (four entrainment, one retest, and one reentrainment). Therefore, ex-

pectation of the subordinate-level precedent should be strong, especially in the Same Speaker condition.

After completing the posttest, participants in the Different Speaker condition completed a recall test, which was intended to examine how well they remembered the different naming precedents the two speakers established for the three filler targets. They were shown pictures of the three filler objects and asked, "What did the first speaker call these objects?" Then, "What did the second speaker call these objects?" We counted the number of pictures for which the listener correctly recalled both terms. (This test was not administered in the Same Speaker condition.)

Next, an oral questionnaire was administered, which is described under "Materials." This was to determine whether listeners believed the cover story and the degree to which they kept the speakers' perspectives separate.

Finally, participants were fully debriefed. We now turn to differences in this general procedure that were introduced in order to manipulate the conversational status of listeners.

Addressee procedure. Participants arrived at the laboratory where they met the experimenter and the female confederate who had recorded the stimuli used in the experiment. We thought that meeting the confederate in person beforehand would make it compelling to listeners that she was actually there when they later listened to her utterances (though in reality, she was not). The female confederate pretended to be a naive participant who was running the experiment for course credit. She arrived 5 min after the appointed time and pretended not to know the experimenter.

The experimenter introduced the idea of a "third" male participant who was "missing." This was to motivate the later change of speaker in the Different Speaker condition. Participants in the Different Speaker condition would hear this person "arrive" later in the experiment (at the beginning of the posttest). In the Same Speaker condition, it simply would appear to listeners that the third person never showed up.

As in Experiment 2, before the experiment began there was a staged "drawing" of instruc-tion sheets that participants believed would determine the roles of "director" and "matcher." Each person then silently read his or her copy of the instruction sheet, which we include as Appendix A. The stated purpose of the experiment was to investigate "the role of feedback and multiple speakers in language comprehension." This justified the separation of the listener into a different room from the director. Listeners believed that the speaker communicated with them through a one-way microphone so that there would be no way for them to provide any feedback.

After the pair finished reading the instructions, the experimenter wondered aloud at the whereabouts of the third participant: "Hmm, the third guy still hasn't shown up. Let's just go on with the experiment anyway. Technically, there are supposed to be two directors, and if he eventually shows up, he can just take over [the confederate's name]'s place."

Next, listeners observed the confederate as she sat down in front of a computer screen and put on a set of headphones and a microphone that was mounted on a headset. Then the experimenter left the confederate behind and led the listener down the hall to the testing room where he or she was seated in front of a computer. At this point, unbeknownst to listeners, the confederate departed and listeners would, for the rest of the experiment, hear her prerecorded utterances.

The computer was connected to a pair of audio speakers through which the participant heard the confederate's utterances. The experimenter wore a headset with a microphone to make it appear that he could verbally interact with the confederate. Staged interactions between the experimenter and confederate occurred at various points during the experiment. These were introduced to make it extremely plausible to listeners that the confederate was really there, in the other room, talking to the experimenter and to them. During these interactions, the experimenter pretended to talk with the confederate. In reality, he was synchronizing his utterances with prerecorded dialogue. These interactions occurred upon arrival to the testing room, before and after the practice trials, and in

the interruption phase just prior to the posttest, where participants in the Different Speaker condition heard a second director take the confederate's place.

After the eyetracker was mounted, listeners completed the practice trials. In a staged interaction that occurred right after the practice trials, listeners heard the experimenter ask if there were any questions, and heard the confederate respond, "Does [participant's name] see the exact same two pictures that I see?" We thought that when listeners heard Speaker A mention their names, they would feel compelled that there was really someone there, talking to *them*. Then the experiment began.

After listeners completed all eight blocks of entrainment trials, and just prior to the posttest, the experiment was interrupted by a staged event. In the Different Speaker condition, listeners heard a staged interaction where Speaker B was heard to "arrive" at the room down the hall where Speaker A gave instructions. Listeners heard someone knocking on the door of the room where Speaker A was believed to be. Speaker A said, "I think there's someone at the door." The experimenter replied, "I wonder if that's the third guy. I'll be right there," and then left the room. While the experimenter was gone, listeners heard a prerecorded interaction between experimenter, confederate, and Speaker B. This established four important beliefs on the part of the listener, namely (1) that Speaker B had never previously met the experimenter nor the first speaker; (2) that he had been absent during the first part of the experiment and thus could not possibly have knowledge of A's descriptions; (3) that, like the listener, he was also a naive participant; and (4) that B was replacing A, who they could hear depart. This part of the procedure required approximately 3 min.

To make the interruption comparable in the Same Speaker condition, this staged interaction was replaced by a 3-min episode where the confederate complained that her computer had crashed. The experimenter left the testing room and listeners heard the experimenter enter A's room, chat informally with A, and restart her computer.

In both Speaker conditions, upon returning to the testing room, the experimenter recalibrated the eyetracking equipment and began the posttest trials.

After completing the experiment, listeners were asked to rate the credibility of the cover story on a scale of 1 to 10, with 1 corresponding to *didn't believe it at all* and 10 to *fully believed it*.

Overhearer procedure. Participants arrived at the laboratory and met the experimenter. They did not meet the confederate. The experimenter told them that the experiment investigated the role of feedback and multiple speakers in how people understand words and that they would be listening to prerecorded utterances from previous participants. They were told to read the instructions (Appendix A) that the participants before them were given. In the Same Speaker condition, the experimenter told the participant that they would hear descriptions from a single speaker because the second speaker did not show up. In the Different Speaker condition, they were told that the second speaker arrived late and that they would not hear him until the end.

They completed the practice trials and heard the same staged interactions, but of course, they believed that these were all things that had happened before. Instead of the experimenter synchronizing his utterances with someone who was not there, the listener simply heard his prerecorded voice in the playback mix. The names that listeners heard the confederate use after the practice trials were not their own, but that of one of the participants in the Addressee condition. The staged events in the interruption phase for the Different/Same Speaker conditions were the same as in the Addressee condition. Unlike the Addressee condition, however, the experimenter remained in the room with the participant for the entire duration.

Materials. Forty-two bitmap images were used in the construction of stimuli. These images depicted everyday objects from a variety of categories, including animals, vehicles, flowers, kitchen utensils, furniture, and so on. The size of the bitmaps was 278×278 pixels, and they were shown on a computer monitor set at a resolution of 1024×768 pixels.

Each display consisted of two images, centered vertically on the computer monitor and spaced equidistantly from the vertical midline on the horizontal axis. Each display was paired with a spoken stimulus to constitute an experimental item.

There were eight sets of items in the experiment, created from combining four different bitmap images from two categories (e.g., car and flower). Each set contained two critical displays that were used in the pre- and posttests, where the subordinate-level precedent for one picture (e.g., "carnation") shared an overlapping phonological onset with the basic-level name for the other (e.g., "car"). The two other images (e.g., "station wagon" and "daisy") were from the same category as the basic-level target and subordinate-level competitor. The basic/subordinate level pairings for all eight sets were "car–carnation," "pitcher–pitbull," "butterfly–butterknife," "tape–tablespoon," "coat–cobra," "clock–clawhammer," "plant–plank," "tent–tennis shoe."

Target images appeared on either the left or right. There was no systematic placement of targets within or across item sets.

We modeled the auditory stimuli after naturally produced names. The female confederate, who later served as Speaker A, participated as speaker in a mock version of the experiment, with the experimenter as addressee. This was before she knew the purpose of the experiment or had ever seen the images. We recorded her voice as she named the target images. The recordings were transcribed to generate a script. Some of the subordinate-level descriptions were changed in the script so that they shared the same onset with the basic-level term. All of the names that were later used in the experiment were rerecorded from the script, and the earlier recordings were discarded.

For the entrainment phase, four different tokens of each name were recorded. To make the speech sound realistic, the linguistic characteristics of each token depended on the speaker's familiarity with the referent. Initial referring expressions included such features as repairs ("the car . . . I mean, the sports car"), hesitations ("um . . . the . . . stained tile of wood"), and a rising final intonation. For example, when referring to a particular dog for the first time in the context of a pit bull, the speaker said, "oh, what kind of dog is that? um . . . golden retriever?". Later expressions from the speaker had a shorter duration, fewer hesitations, and a more confident intonation (e.g., "the golden retriever").

A male native speaker of American English recorded the posttest utterances for Speaker B. To make the acoustic features of the two speakers' utterances as similar as possible, the second speaker modeled his own after those of the first speaker. For each posttest item, the waveforms of the two speakers' utterances were visually compared to ensure that timings of the point of disambiguation were roughly matched. The average point of disambiguation for Speaker A's test utterances was 235 ms, and 230 ms for Speaker B.

The oral questionnaire that was administered at the end of the experiment was as follows:

How strongly would you endorse the following statement(s), on a scale of 1 to 10, with 10 being strongest?

1. I believe that the second speaker had no knowledge of the first speaker's descriptions. (Different Speaker condition only).

2. I believe that I was listening to (a) live speaker(s).

Analysis. Interference in the posttest could manifest itself in two ways: (1) as an increased likelihood of gazing at the competitor image or (2) as a delay in eye movements to the target. For statistical analysis we used a single measure that combines these two possibilities, the *target preference score*. We grouped the samples into 24 ms "bins," computing a preference score for each bin. Each 24 ms corresponded to six samples of eye data (the EyeLink system samples at a rate of 250 Hz, or 1 sample every 4 ms). Each bin was labeled with the median sample number. Thus, bin 12 corresponded to samples spanning 0 and 23 ms.

The preference score for each bin was calculated by subtracting the number of samples that visual activity was directed toward the competitor from the number of samples that the activity was directed toward the target, and then dividing this value by the total number of samples in the

time span. A positive value for a given span (maximum of 1) means that the listener spent more time looking at the target than at the competitor. Correspondingly, a negative value (minimum of −1) means that the listener looked at the competitor more. A value of zero means they looked at them equally often (or not at all).

We analyzed a temporal window spanning, inclusively, bins 226 and 1068. The former number represents the earliest point at which one would expect to find eye movements programmed on the basis of the linguistic signal (after approximately 50 ms of input), given 180 ms for "saccadic overhead" (Matin, Shao, & Boff, 1993). Bin 1068 represents median RT, after which point data is lacking for half of the trials. In those trials where the participant responded before the median RT, the empty cells up to 1068 were filled with ones (representing preference for the target). In other words, for this analysis the final fixation on the target can be regarded as "cumulative."

Results and Discussion

Manipulation check. We first report our manipulation check measures to show that the cover story was effective and that listeners were able to separate the perspectives of the two speakers.

First, results from the postexperiment questionnaire show that the cover story was extremely effective. The participants in the Addressee condition gave the credibility of the cover story an average rating of 9.8, with a mode of 10. The minimum rating for the credibility of the cover story was 9. Furthermore, all of the participants in this condition indicated that they strongly believed that they were listening to a live speaker, providing a mean rating of 9.6 (SD = .712), with a mode of 10 and a minimum rating of 8 (4 participants). This was reliably different from the Overhearer case, where the mean was 1.6 [SD = 1.23, mode = 1, with a maximum rating of 6 from one participant and missing data from one participant; $t(61)$ = 32.15, $p < .0001$].

Listeners in the Different Speaker condition indicated that they very strongly believed that the second speaker did not know the first speaker's precedents. In the Addressee condition, the mean rating was 9.8 (SD = .712,

mode = 10) compared to 9.2 (SD = 1.23, mode = 10) in the Overhearer condition. This difference was marginal, $t(28)$ = 1.68, $p < .10$. The lowest rating obtained was a 7, given by two participants, one in each of the two cover story conditions.

It was clear that listeners kept the perspectives separate, because they were typically able to recall the different names that the two speakers used for the three filler items. On average, they got 2.52 of the three recall items correct (SD = .58, mode = 3; 4 of 32 data points were missing because the experimenter forgot to administer the test). The mean for the Addressee condition was 2.58 (SD = .51, mode = 3), and 2.46 for the Overhearer condition (SD = .64, mode = 3). There was no difference across Listener Status, $t(61)$ = .39. p = .70.

Main results. We submitted the target preference scores to an ANOVA with Test (Pre vs Post), Speaker (Different vs Same), Listener Status (Addressee vs Overhearer), and Bin as factors. We found a main effect of Test, $F1(1,60)$ = 36.57, MSE = 0.7076, $p < .0001$; $F2(1,7)$ = 35.18, MSE = 0.3690, $p < .0001$. Mean target preference was .62 in the pretest compared to .47 in the posttest. The interaction of Bin with Test was significant (after Greenhouse–Geiser correction), indicating that the temporal characteristics of the comprehension process was different in the two test conditions, $F1(35,2100)$ = 13.89, MSE = 0.035, $p < .0001$; $F2(35,245)$ = 3.74, MSE = 0.0651, $p < .05$. This finding suggests that listeners expected speakers to use subordinate-level precedents in the basic-level context of the posttest.

Listener Status did not interact any of the other factors (all $Fs \approx 1$).[6] We therefore collapsed over Listener Status in order to focus on the Speaker manipulation. The mean target preference for the Different Speaker condition was .62 in the pretest and .47 in the posttest. This was identical to the corresponding values in the

---

[6] Test × Listener Status, $F1(1,60)$ = .4, MSE = .7076, $F2(1,7)$ = 2.20, MSE = .0664, $p < .19$; Test × Listener Status × Bin, $F1(35,2100)$ = .42, MSE = .0349, $F2(35,245)$ = .4, MSE = .0184; Test × Listener Status × Speaker, $F1(1,60)$ = .24, MSE = .7076, $F2(1,7)$ = .42, MSE = .1933; Test × Listener Status × Speaker × Bin, $F1(35,2100)$ = .90, MSE = .0349, $F2(35,245)$ = .61, MSE = .0254.

Same Speaker condition. An ANOVA confirmed that there was no difference in interference between these conditions [Test $\times$ Speaker interaction, $F1(1,60) = .42$, $MSE = .7076$; $F2(1,7) = .06$, $MSE = .3716$; Test $\times$ Speaker $\times$ Bin interaction, $F1(35,2100) = .84$, $MSE = .0349$, $F2(35,245) = .59$, $MSE = .0250$]. These results support the partner-independence hypothesis. Listeners attempted to apply linguistic precedents in the posttest simply because they were available, causing interference in selecting the target. Even though listeners in the Different Speaker condition possessed strong evidence against mutual knowledge of the precedents, they were no less likely to use them to interpret speech than listeners in the Same Speaker condition.

Once listeners observed in the posttest that the speaker was using basic-level terms instead of the precedents, they prepared for this event. In both Speaker conditions there was a decrease in interference from the first to the second half of the posttest trials, as manifest by an increase in the target preference score from .42 to .52, $F1(1,60) = 5.81$, $p < .05$; $F2(1,7) = 10.07$, $p < .05$. However, this factor did not interact with the Speaker variable (all $Fs \leq 1$); the means were .41 vs .53 for the Same Speaker condition, .43 vs .52 for the Different Speaker condition.

To more closely determine the time course of precedent use, we undertook a bin-by-bin analysis of the preference scores using a multistage Bonferroni procedure, with an alpha level of .05 for each set of bin-by-bin comparisons. We performed both a participant and an item analysis to determine at what point preference for the target became statistically reliable. We consider the average of the two analyses to represent the most probable real value. We first report results collapsing across the Speaker conditions and then individually compare the two conditions.

In the pretest, where listeners heard basic-level descriptions, they became more likely to look at the target than the competitor at 312 ms. Allowing 180 ms for programming the eye movement to the target (or for deciding *not* to look away from the target), listeners identified the target after only 132 ms of speech input. This is well before the earliest disambiguation point (185 ms) and about 104 ms before the av-

erage disambiguation point (236 ms). In other words, even without precedents, listeners identified referents on the basis of minimal linguistic input. This is consistent with the predictions of word recognition models such as Cohort (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978) or TRACE (McClelland & Elman, 1986).

In the retest, after the speaker had entrained on subordinate-level precedents, listeners heard the speaker continue to use these precedents (e.g., "carnation"), but in a basic-level context (e.g., while viewing the pictures of the car and flower). They were reliably more likely to fixate the target by 444 ms, on the basis of about 264 ms of input. This was just 26 ms after the average disambiguation point of 238 ms (which ranged from 185 to 367. That listeners identified the carnation as the target so quickly suggests that even when basic-level terms are appropriate, listeners are not surprised when speakers continue using precedents. If listeners had been surprised by the precedents, it seems likely that a delay much larger than 26 ms, relative to disambiguation, would have been observed.

Finally, in the posttest, listeners heard the speaker return to the basic-level name ("car") in a basic-level context. If they expect the precedent ("carnation"), then they should experience interference in selecting the target. This is what we found. Preference for the target did not become significant until 528 ms, after listeners had heard about 348 ms of the input. This was well after the average disambiguation point of 232 ms and even the latest disambiguation point of 282 ms. Direct comparison of the posttest with the pretest revealed reliably greater posttest interference from 420 to 744 ms. In other words, participants expected speakers to violate the Maxim of Quantity (Grice, 1975) and continue to use linguistic precedents, even though basic-level terms would have been sufficiently informative in that context.

We now focus our attention on the Speaker manipulation. Figure 4 shows probability plots for these two conditions (Different and Same). Each chart displays fixation probabilities for the target (squares) and the competitor (circles). Additionally, each chart shows probabilities from the pretest (filled in squares and circles)

and the posttest (open squares and circles). The x axis is labeled with bin number. At bin 12 the probabilities do not start at zero because of the free-viewing paradigm.

It can be seen very clearly from the plots that expectation of the precedent on the part of listeners interfered with their interpretation of the basic-level term in both Speaker conditions, ruling out the strong partner-specificity hypothesis. In the Different Speaker pretest, fixations to the target ("car") became reliably more likely than fixations to the competitor ("carnation") at 288 ms. In the posttest, this did not happen until 564 ms, amounting to a delay of 276 ms relative to the pretest. In the Same Speaker condition, target fixations were reliably more likely by 372 ms in the pretest and 576 ms in the posttest. In other words, there was a delay of 204 ms, which is numerically smaller than the delay of 276 in the Different Speaker condition; in other words, this is in the opposite direction from what the strong and weak partner-specificity hypotheses would predict.

To summarize, our results indicate that listeners, like speakers, entrain on linguistic precedents, and expect speakers to overspecify referents. When they heard basic-level terms in the posttest instead of the precedents, they experienced interference in selecting the target, even though the basic-level term would have been sufficiently informative.

Listeners used linguistic precedents because these precedents were available to them, not because they were mutually known. Subordinate-level precedents such as "carnation" did not reliably produce more interference in recognizing the word "car" when it was mutually known than when it was simply available. There was no evidence for either weak or strong partner-specificity.

To conclude, we have shown that listeners' expectations that speakers will use linguistic precedents is strong; in fact, so strong that listeners expect to hear them even when they are overinformative—and surprisingly, even when they have no reason to believe that they are shared with the speaker.

Clearly, this experiment presents the strongest case against the claims of the partner-specificity of linguistic precedents. However, its task might strike some readers as unnatural, which could possibly limit the generality of the results. One way that the task diverges from face-to-face conversation is the lack of interaction between speaker and listener. How might the lack of interaction cause listeners' precedent use to differ from interactive situations?

First, perhaps the only way a precedent can be mutually accepted and added to common ground is through live interaction, as grounding theory suggests (Clark & Brennan, 1991). Note, however, that under this assumption, listeners in Experiment 3 should never use precedents—but clearly, we show that they do. Therefore, it is possible that certain effects of precedent use that have been claimed to support the collaborative model may instead be attributed to low-level availability effects. For instance, Brennan and Clark (1996) present the finding that speakers overspecify referents as evidence for the existence of conceptual pacts. However, as they note, there are factors other than conceptual pacts, such as frequency or recency, that might cause speakers to continue to use a pact. It is not clear whether the overspecification they observed was due to the existence of conceptual pacts or to these other factors because the interactivity of the situation was not manipulated. For instance, it is possible that speakers would have overspecified referents even without interaction simply because of the availability of the precedents.

A more troubling criticism would be that the absence of true interaction prevented listeners from sufficiently associating Speaker A's precedents with that speaker to yield partner-specificity effects. We acknowledge this as a possible limitation of this experiment, but it should be kept in mind that we did not find partner-specific effects in the face-to-face situation of Experiment 2, where listeners interacted with a confederate. The common view that partner-specific effects in processing single utterances are most likely to be found in live interaction is simply an a priori assumption with no empirical support. In fact, we suggest that there are even better a priori reasons to expect that live interaction is where one is *least* likely to find partner-specific effects because the presence of immediate, multimodal feedback permits greater egocentrism (Barr, 1999).
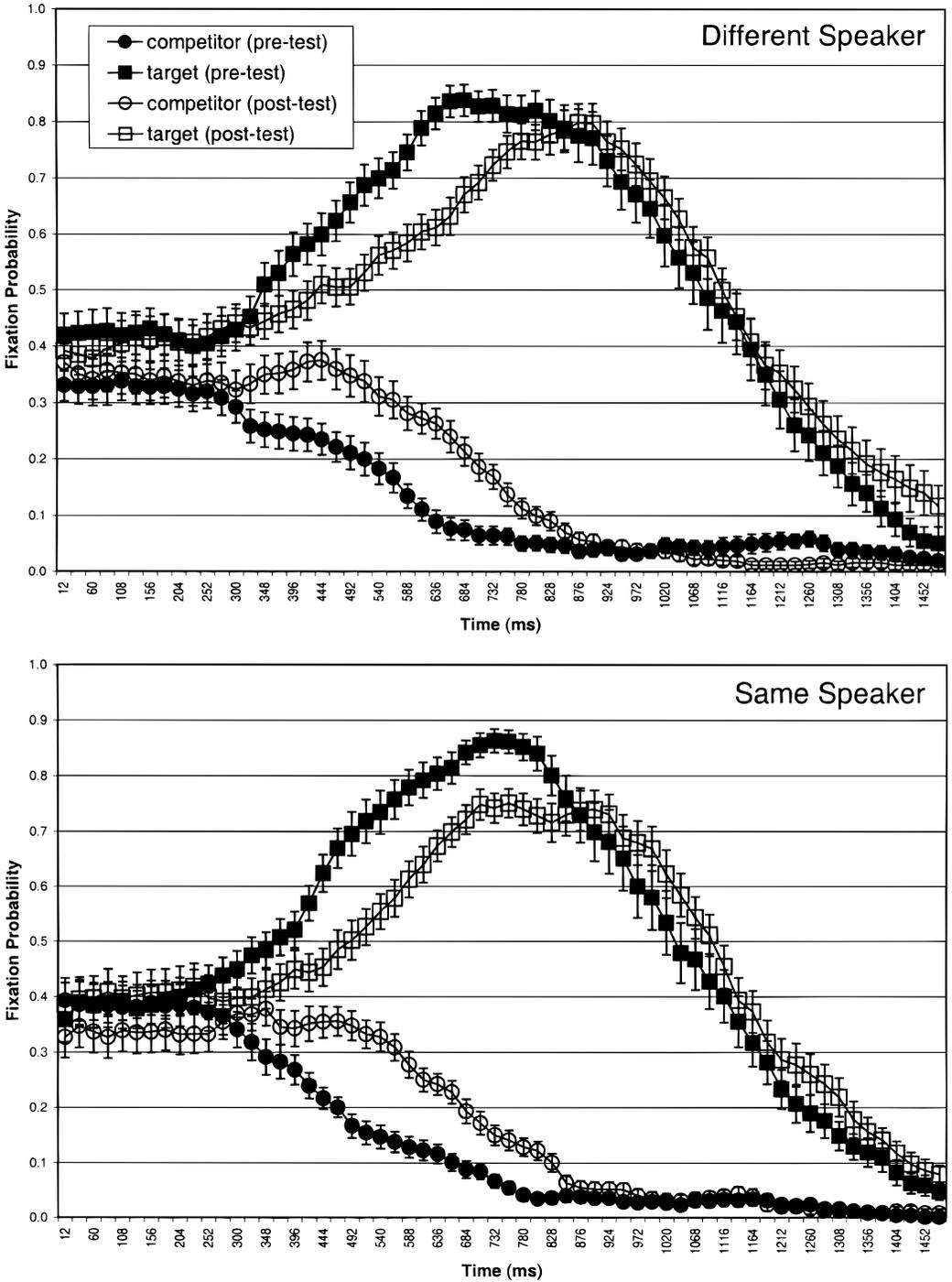
FIG. 4. Fixation Probability Plots in Experiment 3. Error bars represent the standard error of the mean for each measure.

Another reason the task of Experiment 3 may strike some as unnatural is that the naming task is not embedded in any broader conversational task. This too could fail to generalize in more than one way. On the one hand, it may heighten attention to the task itself and allow the development of processing strategies that would not arise in a more conversational task. On the other hand, in a conversational task, attention is distributed over many more processes, perhaps making it even more difficult for listeners to keep track of who knows what. In other words, the task of Experiment 3 could have potentially *overestimated* partner-specificity effects because there listeners could concentrate more attention who used what name, given that they were not distracted by other things. Finally, note that we failed to find partner-specific precedent effects in Experiment 2, where the naming was embedded in the conversational task of rearranging objects.

GENERAL DISCUSSION

Our experiments demonstrate that the standardization that occurs in language production has a counterpart in comprehension. Linguistic precedents enable listeners to quickly access referential meanings and can override the assumption that speakers will be optimally informative in their referential descriptions. Listeners expect speakers to follow precedents, whether the precedent has been used only once or many times. Of course, the strength of a precedent will depend on its frequency and recency, though our experiments do not examine these factors.

In the first experiment, the benefit was much larger for unconventional than conventional referents, clearly implying that it was *linguistic*: it depended on an established relation between the label and the object. The second experiment found benefits at second mention, while the third experiment found benefits due to entrainment on the precedent. Both Experiments 2 and 3 supported partner independence over the partner-specificity hypotheses: listeners were just as likely to use precedents with a new speaker as they were with the one who established them. This led to an equal comprehension benefit in Ex-

periment 2 in the two Speaker conditions. In Experiment 3, listeners' expectation of precedents impaired comprehension of basic-level terms, but listeners were no less likely to expect them with a new speaker than with an old one. Not even weak partner specificity was observed, suggesting that listeners use precedents because they are available, not because they are mutually known.

As a point for future research, Experiment 3 suggests interesting possibilities for models of word recognition. Consistent with Cohort and TRACE models, we found that in the pretest listeners could access referential meanings on the basis of limited phonological input. In the retest, listeners accessed subordinate-level precedents extremely quickly, even though in the context of the display, the basic-level name would have been appropriate. Models of word recognition might easily incorporate linguistic precedents as a boost in the baseline activation of words that correspond to precedents.

One criticism of our studies might be that they require people to represent what others do *not* know, something that has been characterized as computationally intractable (Polichak & Gerrig, 1998). This claim is not very compelling because people can and very often do represent the fact that others do not know something that they themselves know; in fact, this state of affairs is what often stimulates people to communicate in the first place. Even if we grant this criticism, our studies do not require a listener to represent *everything* that a speaker does not know, but simply to ignore the precedents that were established by the previous speaker. For instance, in Experiment 3, during the one second that listeners were provided to scan the display before hearing the name, they could have prepared for the upcoming basic-level name from the speaker by inhibiting the precedent and activating the basic-level term. In fact, listeners showed that they *could* inhibit precedents, as reported above in the decrease in interference from the first to the second half of the posttest. Yet even while listeners in the Different Speaker condition had greater justification than Same Speaker listeners for both inhibiting Speaker A's precedents and anticipating basic-level terms, they were not more likely to do so.

Besides, this criticism misses the point that the design we use in our studies is essential to demonstrating that a process distinguishes between what is and what is not mutually known. Studies that offer evidence for the use of mutual knowledge typically demonstrate that language users rely on information that is mutual (e.g., Clark, Schreuder, & Buttrick, 1983; Greene, Gerrig, Mckoon, & Ratcliff, 1994). But as Keysar (1997) demonstrated, these studies confound mutuality with information available to the self. Every piece of information that is mutually known to A and B is also simply known to A. Therefore, to support the argument that comprehension relies on mutual knowledge, it is not sufficient to show that mutually known information was used—one must also show that information that was not mutual was *not* used. This would demonstrate that the process in question distinguishes between mutual and available information.

It is interesting that the comprehension system is not designed to effectively use mutual knowledge, which is a potentially valuable source of information. Clark and Carlson (1981) proposed that mutual knowledge can potentially limit the scope of contextual information that the comprehension system must consider. This would prevent listeners from making systematic mistakes. It may be, however, that this benefit in information reduction and increased success is outweighed by the difficulty of using meta-knowledge during real-time language processing. In our view, listeners' use of precedents that are outside mutual knowledge is symptomatic of a language processing system that is designed to use available information to settle matters of referential ambiguity quickly and efficiently so that it can keep up with the rapid influx of linguistic information. In other words, it is a system that is designed in response to the compromises of performance in the real world, not to the exacting standards of pragmatic theories.

To preempt a possible misunderstanding of our position, we point out that although listeners in the current experiments did not use the identity of the speaker to guide the identification of the referent, this does not imply that they were not keeping track of this information. It is obvious that language users can and do represent who said what. For instance, participants in Experiment 3 were typically able to recall the conceptual precedents that the two different speakers had established. The important point is that they simply did not *use* this knowledge to identify intended referents. There may be other domains of language use where this knowledge could be effectively used.

It may be noted that our studies seem to conflict with research on precedent use in language production, where some researchers have made claims for partner specificity (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986). For instance, Brennan and Clark (1996) made strong claims for the partner specificity of conceptual pacts: "In the historical models so far, speakers choose their wording regardless of whom they last spoke to. But according to partner specificity, they do so for the specific addressees they are now talking to" (p. 1484). But Brennan and Clark did not find this. Instead, they state that, "the references in our task emerged not from solitary choices on the part of the director, but from an interactive process by both director and matcher" (p. 1491). In other words, they appeared to show that "*dyads negotiate*" references over prolonged interactions, not that "speakers choose" them in producing single utterances. What this means is that when directors started talking to new addressees, they used the precedents they had established with the old matcher. They then received feedback from their addressee, who were surprised by the overly specific terms, and because of this feedback they changed their referring expressions. Such pattern of results concerning speakers' behavior is perfectly consistent with our results concerning addressees' behavior. In neither case need one assume that a partner-specific pact was established between matcher and director. Instead, it is sufficient to assume that each of them entrained on precedents ("loafer" and "carnation") and later attempted to continue using them with new partners, but were rebuffed. Such an account is consistent with the more general anchoring and adjustment model of Keysar and Barr (in press), where speakers and listeners use mutual knowledge only to diagnose and correct coordination problems.

Evidence for the partner-specific use of precedents in language production is lacking. Other researchers in the area have found that certain accommodations that speakers appear to make for addressees are actually due to low-level effects such as availability or priming (Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, & Newlands, 2000; Brown & Dell, 1987; Ferreira & Dell, 2000). More studies are needed, especially like the above-mentioned ones, which overcome the confound discussed by Keysar (1997).

The results from our three experiments help put Henry's egocentrism in context. Henry retrieves the reports from the detective agency as the referent of "the reports" because he anchored his understanding in the linguistic precedent that Bendrix had established immediately prior to the exchange. This led him to the incongruous conclusion that the newcomer had knowledge of an extremely private and sensitive matter. He knows this is implausible and attempts to adjust away from the anchor, but his wife's possible infidelity lies heavily on his mind and is not easily ignored. Because of this, he asks the speaker "What reports?" even though there is only one referent that is uniquely defined by their mutual knowledge. In short, the nature of the comprehension system makes us all prone to misunderstandings such as Henry's—even when our spouses are not cheating on us.

APPENDIX

Instruction Sheet for Experiment 3

INSTRUCTIONS FOR: MATCHER

This experiment investigates the role of feedback and multiple speakers in language comprehension. We are interested in how people understand descriptions from different people, under conditions where feedback is either present or absent.

You are in the **FEEDBACK ABSENT** condition.

You have drawn the role of the **MATCHER**. The other two participants will take turns being the **DIRECTOR**. To be completely sure that directors are unable to get any verbal or nonverbal feedback from you, we will put you (the matcher) in a separate room down the hall (Room 1428).

You will be performing the task on a computer. Your computer (which is located in 1428) is connected to the directors' computer (the one here, in 1410) through a local Intranet. The task is very simple. Two pictures will appear simultaneously on your screen and on the directors' screen. One of these pictures is the "target" picture. A director will name that picture for you. Based on that person's description, you will select the picture that you think is the target. Indicate your selection by clicking on the picture using the computer mouse. Because you are in the **FEEDBACK ABSENT** condition, you will not be able to give the directors feedback or ask them for help. Furthermore, the directors will not know if their descriptions are successful. Thus, if you are unable to determine the target picture based on the description, just make your best guess.

Each director wears a private set of headphones. Before the pictures appear, one of the two directors will hear a voice in the headphone say "right" or "left". This tells that director that: (1) it is her/his turn to name the target; and (2) that the target will appear on the corresponding (right or left) side of the screen. Directors are NOT given names for targets in the headphones—they only know their locations. Therefore, they must come up with their own way of naming them. We will be recording their descriptions digitally onto the computer's hard disk.

As you perform the task, you will be wearing a headband equipped with two cameras that film your eyes. Later, we will analyze this data with respect to the descriptions that you heard. This allows us to understand how the eyes change in response to spoken language. The experimenter will explain the apparatus in more detail once you arrive in Room 1428.

Before the experiment begins, you and the two directors will participate in three practice trials together to make sure that everyone properly understands the task.

Please feel free to ask the experimenter any questions at this time.

REFERENCES

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, **42**, 1–22.

Barr, D. J. (1999). *A theory of dynamic coordination for conversational interaction*. Ph.D. thesis, The University of Chicago.

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

Brown, R. (1958). How shall a thing be called? *Psychological Review*, **65**, 14–21.

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, **19**, 441–472.

Clark, H. H. (1982). The relevance of common ground: Comments on Sperber and Wilson's paper. In N. Smith

(Ed.), *Mutual knowledge* (pp. 124–127). London: Academic Press.

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

Clark, H. H., & Carlson, T. B. (1981). Context for comprehension. In J. Long & A. Baddeley (Eds.), *Attention and Performance* (Vol. IX, pp. 313–330). Hillsdale, NJ: Erlbaum.

Clark, H. H., & Carlson, T. B. (1982). Speech acts and hearers beliefs. In N. V. Smith (Ed.), *Mutual knowledge* (pp. 1–37). London: Academic Press.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshe, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, UK: Cambridge Univ. Press.

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

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, **22**, 245–258.

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

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, **24**, 121–135.

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

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, **27**, 181–218.

Greene, S. B., Gerrig, R. J., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language*, **33**, 511–526.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.

Hanna, J. E., Trueswell, J. C., Tanenhaus, M. K., & Novick, J. M. (1997) *Consulting common ground during referential interpretation*. Poster presented at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, PA.

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

Keysar, B. (1997). Unconfounding common ground. *Discourse Processes*, **24**, 253–270.

Keysar, B., & Barr, D. J. (in press). Self anchoring in conversation: Why language users don't do what they "should." In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *The psychology of judgment: Heuristics and biases*. Cambridge, UK: Cambridge Univ. Press.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The use of mutual knowledge for error correction. *Psychological Science,* **11,** 32–38.

Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, **39**, 1–20.

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

Krauss, R. M., & Weinheimer (1964). Changes in reference phrases as a function of frequency of usage in social interaction. *Psychonomic Science*, **77**, 622–626.

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

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition*, **25**, 71–102.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, **10**, 29–63.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception & Psychophysics*, **53**, 372–380.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1–86.

Polichak, J. W., & Gerrig, R. J. (1998). Common ground and everyday language use: Comments on Horton and Keysar (1996). *Cognition*, **66**, 183–189.

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

Sperber, D., & Wilson, D. (1982). Mutual knowledge and relevance in theories of comprehension. In N. Smith (Ed.), *Mutual knowledge* (pp. 61–87). London: Academic Press.