Conceptual Pacts and Lexical Choice in Conversation

Susan E. Brennan
State University of New York at Stony Brook

Herbert H. Clark
Stanford University

When people in conversation refer repeatedly to the same object, they come to use the same terms. This phenomenon, called *lexical entrainment*, has several possible explanations. *Ahistorical* accounts appeal only to the informativeness and availability of terms and to the current salience of the object's features. *Historical* accounts appeal in addition to the recency and frequency of past references and to partner-specific conceptualizations of the object that people achieve interactively. Evidence from 3 experiments favors a historical account and suggests that when speakers refer to an object, they are proposing a conceptualization of it, a proposal their addressees may or may not agree to. Once they do establish a shared conceptualization, a *conceptual pact*, they appeal to it in later references even when they could use simpler references. Over time, speakers simplify conceptual pacts and, when necessary, abandon them for new conceptualizations.

When speakers refer to an object, as with *the loafer*, one of their goals is to get their addressees to identify the object. Current theories of reference differ on how they manage that. By some theories, speakers design each referring expression with enough information, but no more than enough, to distinguish the referent from its surroundings. By other theories, speakers may include more information than necessary. Both of these types of theories are ahistorical: They assume that speakers design references without regard to past or potential interaction with addressees. In this article, we argue that speakers exploit both past and potential interactions, and we test several alternatives for how they do that.

Labels reflect conceptualizations. According to the principle of contrast (E. V. Clark, 1987), all forms of language contrast in meaning, and there are no true synonyms (see also Bolinger, 1977; H. H. Clark & Clark, 1979; Levelt, 1989). So when speakers refer to *the loafer*, they are conceptualizing an object, for current purposes, as a loafer, not merely as a piece of clothing, shoe, or casual shoe (Brown, 1958a, 1958b). How do speakers choose and mark the particular conceptualization they use? Let us first consider three ahistorical factors: informativeness, lexical availability, and perceptual salience.

The best known proposal is that speakers simply design their referring expressions to distinguish a referent from a set of

alternatives (Brown, 1958b; Olson, 1970): They choose the most concise expression that will enable their addressees to pick out the referent uniquely. By Grice's (1975) maxim of quantity, for example, speakers try to provide enough information ("Make your contribution as informative as is required") but without providing too much ("Do not make your contribution more informative than is required"; Grice, 1975, p. 26; see also Horn, 1984; Levinson, 1983). So if they want to refer to a brown loafer in a set consisting of the loafer, a sneaker, and a high-heeled shoe, they should not use *the shoe*, which is not informative enough, or *the brown loafer*, which is more informative than required, but rather *the loafer*.

Informativeness, however, may be overridden by other factors. One such factor is lexical availability—how easy it is to conceptualize a thing in one way and to retrieve and produce a corresponding label. For many objects, the most available labels are basic-level nouns—*dog*, for example, as opposed to *animal* or *terrier* (Cruse, 1977; Murphy & Smith, 1982; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). When the alternatives consist of a terrier, a bicycle, and a hammer, by Grice's (1975) maxim of quantity speakers should refer to the terrier as *the animal*. In fact, they are more likely to choose *dog* even though it is more informative than required. Another factor that can override informativeness is perceptual salience. Sometimes speakers describe what is salient about an object (e.g., *small black terrier*) rather than just those features that will distinguish it from its neighbors (Ford & Olson, 1975; Mangold & Pobel, 1988). When the set of alternative objects is large, speakers may be overinformative if they begin their references before they have taken in the entire array (Deutsch & Pechmann, 1982; Pechmann, 1989). According to one account of lexical choice (Cruse, 1977), the most common term that is also informative enough should be considered to be unmarked; all other terms are marked.

Even in light of informativeness, availability, and salience, there is still a great deal of variability in speakers' lexical choices. The problem is that most objects can be conceptualized in indefinitely many ways. In one study, when people had to label an action to be carried out by a computer, the likelihood that any two people would use the same label,

Susan E. Brennan, Department of Psychology, State University of New York at Stony Brook; Herbert H. Clark, Department of Psychology, Stanford University.

Correspondence concerning this article should be addressed to Susan E. Brennan, Department of Psychology, State University of New York, Stony Brook, New York 11794-2500. Electronic mail may be sent via Internet to susan.brennan@sunysb.edu.

conceptualizing the action in the same way, was only 7–18% (Furnas, Landauer, Gomez, & Dumais, 1987). In another study, pairs of people in conversation referred to the same abstract geometric form variously as "the rice bag," "the whale," "the complacent one," "the stretched-out stop sign," and "the baby in a straitjacket" (Schober & Clark, 1989, p. 217). We would like to understand variability and consistency in lexical choices for at least two reasons. First, predicting lexical choices should improve speech recognition by machines (Brennan, Ries, Rubman, & Lee, 1996). Second, if lexical choices reflect conceptual coordination in conversation (H. H. Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987), then they should tell us not only about individual processes of language use, but also about distributed ones.

Historical Models of Referring

The factors described so far account for "one-shot" referring; the tacit assumption is that there is no past or future interaction between speakers and addressees. In conversation— the primary site of language use—it is known that people design their referring expressions on the basis of their past references to the same objects (Carroll, 1980; H. H. Clark & Wilkes-Gibbs, 1986; Krauss, 1987; Krauss & Weinheimer, 1964, 1966; Schober & Clark, 1989). With repeated references to objects, they tend to reuse the same terms as they coordinate their perspectives, a phenomenon that Garrod and Anderson (1987) have labeled *lexical entrainment*.[1]

Historical models differ from ahistorical models in their predictions about lexical choices. Suppose there are two sets of 12 cards, Set A and Set B, with a picture of a common object on each card. Both sets include the same loafer; this is the only shoe in Set A, whereas Set B has a high-heeled shoe and a sneaker as well. Now suppose two people, a director and a matcher, are asked to talk about these objects one set at a time. And in doing that, the director refers first to all the objects in Set A, then to all the objects in Set B, and then again to all the objects in Set A. We call these the A, B, and C trials. For simplicity's sake, we refer to the director as female and the matcher as male in examples.

The issue is how the director will refer to the loafer. In the A trials (with Card Set A), she will probably refer to it as *the shoe*, using the highly available basic-level term. In the B trials (with Card Set B), she should be more specific, using a phrase such as *the loafer*, to distinguish it from the other two shoes. What will she do in the C trials (with Card Set A again)? In ahistorical models such as Cruse's (1977), she should use *the shoe* because it is both informative enough and highly available. But in historical models, the director might retain *the loafer* even though it is overinformative. To explain why she might do that, let us consider four factors that define increasingly specific models: recency, frequency of use, provisionality, and *partner specificity*.

Recency

With the first factor, recency, people in conversation expect to conceptualize an object the same way they conceptualized it on their last successful reference. This factor is embodied in

Garrod and Anderson's (1987) output/input coordination principle, which holds that people formulate their current utterance according to the same model and semantic rules used to interpret their partner's most recent utterance. This accounts for how they come to use the same terms and share an underlying conceptualization (Garrod & Anderson, 1987). The output/input coordination principle has several a priori advantages. First, because it is strictly local, all speakers have to do is keep track of the last description. That should simplify their job in forming a reference. Second, the principle does not assume that speakers and their partners necessarily form complicated models of each other or beliefs about their shared knowledge—their mutual beliefs—and that also should simplify the process of referring.

Recency, however, does not tell the whole story. It does not account for how people choose their terms for a brand-new referent. If their initial choice is based on informativeness, availability, and salience, then recency is a plausible basis only for re-referring. Second, in a historical account based on recency alone, repeated references to an object would seem to result in rigid conformity, for "once co-ordination has been achieved, no modification can be introduced without violating the principle of local consistency" (Garrod & Anderson, 1987, p. 209). But as Garrod and his colleagues noted (Garrod & Anderson, 1987; Garrod & Clark, 1993; Garrod & Doherty, 1994), description schemes often continue to evolve, which prompted them to offer several modifications to the output/input coordination principle. The first modification assumed that "one participant, the leader, [is] the arbiter over the established scheme, while the other participant, the follower, [has] to abide rigidly by the output/input principle" (Garrod & Anderson, 1987, p. 209). The second assumed that speakers override the principle when they need to repair a referring expression that has been misunderstood (Garrod & Clark, 1993). And the third assumed that the principle can be overridden by situational salience (Garrod & Doherty, 1994). Even with these modifications, the output/input coordination principle predicts that directors in our C trials should retain the most recent successful reference to an object. It should not matter how they conceptualized the object before that.

Frequency of Use

Lexical choices should also be affected by speakers' memory for a conceptualization. The more often people appeal to a particular conceptualization, the more durable its memory representation should be, as Garrod and Doherty (1994) pointed out.[2] We call this frequency of use. A model based on both frequency of use and recency predicts one additional phenomenon over a model based on recency alone. Consider the situation described earlier in which a director and a

---

[1] A related correspondence effect, the tendency of speakers to repeat syntactic form and lexical items from a question in its answer, was investigated by Levelt and Kelter (1982).

[2] Although Garrod and Doherty (1994) considered the durability of representations in their account of how groups evolve global conventions, they did not investigate this as a factor in local, pairwise coordination.

matcher discuss a set of B cards in which the loafer is not the only shoe and then, in C trials, discuss a set of cards in which it is the only shoe. The more often the director has referred to the loafer in Set B, the more firmly that conceptualization has been established, so the more likely she should be to retain *the loafer* in the C trials. Likewise, if they had discussed the unique A cards several times before discussing the nonunique B cards, then upon returning to the unique cards in the C trials they should be more likely to return to using *the shoe* than if they had never discussed the unique cards before. Frequency of use takes into account how strongly and easily activated a conceptualization is, whereas recency considers only the conceptualization used on the last reference.

Recency and frequency of use are not the only factors in re-referring; if they were, speakers should use the same terms in talking to responsive partners as in talking into a tape recorder, and they do not. With partners, but not with tape recorders, repeated references tend to become shorter and more efficient (Krauss & Weinheimer, 1964, 1966; Kraut, Lewis, & Swezey, 1982; Schober, 1993). In a more complete historical model, then, references are not only based on precedent but are also adapted to changing circumstances. They are modified in response to a partner's feedback and shortened as precedents become more firmly established. We consider this adaptability with the next two factors—provisionality and partner specificity.

Provisionality

According to the collaborative view of language use (Brennan, 1990; H. H. Clark & Schaefer, 1987, 1989; H. H. Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989), establishing a precedent in conversation requires interaction. When speakers present a reference, they do so only provisionally, and they then work with their addressees to establish that it has been understood. When speakers first refer to an object as *the loafer*, they are proposing to their addressees that it be conceptualized as a loafer. The addressees can ratify the proposal ("okay"), modify it ("you mean the man's shoe?"), or solicit another proposal ("which one?") in the process of grounding that reference (see H. H. Clark & Brennan, 1991). The conceptualization they settle on may be quite different from the one originally proposed.

With the process of grounding, lexical choice can be based on precedent and still be adaptable to the changing circumstances. When speakers refer to an object a second time, they mark a conceptualization as identical when they choose the verbatim terms (retaining *the loafer*), as related when they choose related terms (moving from *a kinda preppy pennyloafer* to *the pennyloafer*), but as new when they adopt new terms (moving from *the loafer* to *the shoe*). If speakers want to minimize grounding on a repeated reference, they should rely on the conceptualization they had agreed to before, marking this with identical or related terms. And if speakers want to modify a shared conceptualization slightly, to make referring more efficient, or even dramatically, to cope with changing informational needs, they can rely on the evidence their partners provide in the grounding process to know whether proposed terms are understood and accepted.

Grounding leads to two predictions. First, the less certain speakers are that their partners will accept a conceptualization, the more likely they should be to mark their reference as provisional with hedges such as *sort of* and *kind of*. After a conceptualization is accepted, hedges should disappear. Second, once two partners agree on a conceptualization, it should not matter on the next reference which one proposed or last appealed to it. There is no need for the arbiter in Garrod and Anderson's (1987) scheme.

Partner Specificity

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. The idea is that when speakers and addressees ground a reference, they are creating a *conceptual pact*, a temporary agreement about how the referent is to be conceptualized. So when the same speakers face new addressees, they have to establish new conceptual pacts, and these may not be the same as those established with previous addressees. In the task described earlier, directors should retain the term established on the B trials, say, *the loafer*, when they continue with the same matchers on the C trials. But with new matchers on the C trials, they should return to the common, informative enough basic-level term *the shoe*. That is, when directors face new matchers, they cannot rely on previous pacts, which do not exist. They have to establish new ones.

In the experiments that follow, we consider these four factors cumulatively. Pairs of people conversed in Trials A, B, and C, as we described earlier. The goal of Experiment 1 was to compare an ahistorical model of reference based on informativeness to historical models based on recency alone and recency and frequency of use together. In Experiment 2, we replicated Experiment 1 but with director and matcher alternating roles. By giving both partners equal opportunity to refer, we tested whether a conceptualization is as effective a precedent when introduced by one's partner as when introduced by oneself. We also examined whether speakers mark certain conceptualizations as provisional with hedges such as *kind of*. In Experiment 3, we tested a further feature of conceptual pacts, namely that speakers' choice of terms should depend on conceptual pacts established with specific partners.

The strongest historical model we have proposed leads to four predictions: (a) In the right context, speakers should rely on conceptual precedents even when to do so would be overinformative. (b) Speakers should rely on conceptual precedents more often the more firmly they are established. (c) References are provisional; even after a precedent has been firmly established, it can be changed or abandoned in the right context. (d) Speakers establish conceptual pacts with particular addressees.

Experiment 1

In Experiment 1, pairs of people did zero, one, or four A trials of a matching task using a card set (A) in which each target was unique in its basic-level category; then they did either one or four B trials using another card set (B) that

contained the same targets as before, in addition to pictures of objects belonging to the same basic-level categories as the targets; and finally, they did four more C trials using the original card set (A).

Method

Participants. Forty-eight Stanford University undergraduate students (29 women, 19 men) volunteered as part of their participation in an introductory psychology class. Students were paired on the basis of availability.

Materials. Two identical copies of four sets of picture cards were prepared. Each set of picture cards contained 12 cards consisting of six color pictures of common, everyday objects cut from catalogs and six black and white abstract geometric figures or tangrams. Each picture was glued to a 6-in. × 4-in. index card and laminated. The common objects were chosen so that they could be named easily with basic-level terms. The tangram cards (which require much more effort to describe than the common objects) were added after pilot trials, in order to distract participants from focusing exclusively on the changing figure–ground relationships of the common-object cards and guessing the experiment's purpose.

Two of the six common objects in each card set were designated target items; Sets 1A and 1B contained the same target cards, a shoe (loafer) and a dog (retriever), as did Sets 2A and 2B, a car (hatchback) and a fish (rainbow trout). Targets were considered to be "figure" and the other four common objects in each set were considered to be "ground." In Sets 1A and 2A, each of the four ground items belonged to its own distinct basic-level category. In Set 1A these items were a clock, a bee, glasses, and a sock; in Set 2A, there were a toy, a horn, a lamp, and shorts. Sets 1B and 2B each differed from its corresponding A set only by the four ground items; the ground items in the B sets belonged to the same basic-level categories as the figure items. So Set 1B had two additional shoes (a sneaker and a high heel) and two additional dogs (a Scottish terrier and a cocker spaniel); Set 2B had two additional cars (a sedan, side view, and a red car, front view) and two additional fish (an angel fish and a red and white fish), instead of unique ground items. The pattern of shared cards was the same for the tangram distracters in the sets: For each card set, there were two tangrams that were members of both the A and the B sets, with four tangrams that changed from Set 1A to Set 1B and from Set 2A to Set 2B. In short, the target object pictures in the A sets were all unique in their basic-level categories, while in the B sets they were not.

Procedure. In each pair, one person was assigned to the role of director and the other was assigned to the role of matcher. Each person was given a set of 12 picture cards identical to his or her partner's set. The cards were arranged by the experimenter in a frame on the table directly in front of each person. The director and matcher were seated facing one another at a long table bisected by a barrier that prevented them from seeing each other or each other's cards. Their task was to get the matcher's set of cards arranged in the same order in the frame as the director's set, proceeding from left to right, from the top row to the bottom row. They were told that it was very important to be accurate but to move right along, as they were being timed. They were also told that they could talk to each other as much as they wanted and that their conversation would be audiotaped.

Each of the 24 pairs of students was randomly assigned to one of six groups, which determined the particular sequence of A, B, and C trials they were to do. In A trials, pairs matched an A card set and in B trials, a B card set; in C trials, they reused the A card set. Eight pairs did four A trials, 8 did one A trial, and 8 did none. As for B trials, 12 pairs did four B trials and 12 pairs did one. All 24 pairs did four C trials. After finishing the sequence of A, B, and C trials with one of the card sets (either Set 1 or 2), each pair repeated an ABC sequence with the other

card set. For the second ABC sequence, they did the same number of A trials as with the first card set. Then, if they had done four B trials with Card Set 1, they did one B trial with Card Set 2, and vice versa. So the number of A trials varied between pairs and the number of B trials varied within pairs. Within the groups, Card Sets 1 and 2 were counterbalanced for order (2 of the 4 pairs in each group used Card Set 1 first, and 2 used Card Set 2 first).

This design enabled us to see whether speakers would make their referring expressions for the target figures more informative when the ground changed from the last A (unique) trial to the first B (nonunique) trial and less informative when the ground changed from the last B (nonunique) trial to the first C (unique) trial. In addition, it enabled us to compare the effects of the number of A trials (zero, one, or four) between subjects and the effects of the number of B trials (one or four) within subjects. This design and the sequences of trials are summarized in Table 1.

After each trial, the experimenter removed each set of cards from its frame, shuffled it, and replaced it in a new order, taking care that each picture was right side up for each partner. No target was allowed to appear in the final location in the director's frame, to prevent people from using phrases like *the last one.* After the A trials, the experimenter formed the B card sets by taking the cards to a nearby table (in view of the participants), turning them over, removing the four pictures of common objects and four tangrams that were not targets (the ground items), and replacing them with four pictures that shared basic-level categories with the targets (two per target) and four new tangrams. After the B trials, the experimenter replaced the nonunique ground items with the original ground items. Markings on the backs of the cards (unseen by the participants during the matching task) enabled the experimenter to accomplish this quickly and accurately. Participants were not given any feedback from the experimenter about whether matches were correct. During each trial, the experimenter wrote down all of the director's referring expressions for each of the four target picture cards. Twenty-three of the 24 experimental sessions were audiotape-recorded (1 was not, because of equipment failure).

In earlier pilot trials using only pictures of common objects, the change of figure–ground pattern from A to B and from B to C card sets was extremely salient, and several participants had concluded that their response to figure–ground patterns was what we were studying. The tangrams proved to be effective distractors; in debriefing, none of the 24 pairs in Experiment 1 appeared to have focused on the figure–ground patterns of the picture cards. Most suspected that we were comparing the tangrams and the pictures of common objects.

Analysis and coding. All references to target items in the 23 audiotaped experimental sessions were transcribed later by a research assistant naive to the hypotheses of the experiment. For the 23 audiotaped sessions, 97% of the target references transcribed were identical to those written down in situ. The transcribed references

Table 1
Design of Experiments 1 and 2: Number of Matching Trials With a Particular Card Set

| | Card sets | | |
	A (unique)	B (nonunique)	C (unique, same as A)
Group			
1	0	1, 4	4
2	0	4, 1	4
3	1	1, 4	4
4	1	4, 1	4
5	4	1, 4	4
6	4	4, 1	4

Note. For instance, "1, 4" means that with the first card set, a pair did 1 B trial, whereas with the second card set, that pair did 4 B trials.

from the 23 audiotaped sessions and the written ones from the 1 remaining session were coded for the analyses that follow.

As a measure of lexical entrainment, we considered speakers' consistency in lexical choices over pairs of successive trials. To do this, we computed sequential probabilities that the terms chosen in one trial were the same as the terms chosen in the next trial, within a director–matcher pair. The criterion for equivalence used here was that the referring expressions in both trials had to include all the same content words (so *the fish with the curved tail* was counted as equal to *the fish with its curved tail*, but *the pennyloafer* was not counted as equal to *the pennyloafer shoe*, nor was *the red dog with its mouth open* counted as equal to *the dog with its mouth open*). Changes in the order of terms (e.g., *the man's shoe* to *the shoe, man's*) were counted as equivalent; hedges were ignored (so *red shoe* was counted as equal to *reddish shoe* and *like a loafer* was counted as equal to *loafer*). The sequential probabilities of particular interest were for the last two A, B, and C trials and the pairs of trials that spanned a change of card set (the last A trial, A1 or A4, along with the first B trial, B1, and the last B trial, B1 or B4, along with the first C trial, C1). Equivalence coding was done by two independent coders; 99% of their coding decisions were the same.

Results and Discussion

First we compared the predictions from historical and ahistorical models about informativeness. Indeed, speakers were often more informative than they needed to be in their references, and so informativeness alone does not account for their choices. Then we examined lexical variability and entrainment in the A, B, and C trials to test for the effects of recency and frequency of use.

Informativeness. The base rate for using unadorned basic-level terms for the target items in the first A trial (A1) was high: 77% (75, 75, 94, and 63% for *shoe, dog, car,* and *fish,* respectively). When the reference in A1 was not an unadorned basic-level term, it consisted of or included a lexicalized subordinate term, such as *loafer* (20% of the time) or more rarely, a longer descriptive phrase that included the basic-level term, such as *dress shoe* (3% of the time).

By all of the models of reference we have considered, directors should have followed the first part of Grice's (1975) maxim of quantity (be as informative as required) and they did. In B1, when there were three shoes, three dogs, three cars, and three fish, directors chose terms that were more informative than the ones in their last A trial 89% of the time and more informative than basic-level terms 95% of the time. In the remaining 5% of cases, directors presented the same basic-level terms in B1 as in the last A trial, and matchers prompted them for more information. Performance for matching target cards was at ceiling: One matcher made one error and repaired it later in the trial.

Although speakers seemed to heed the first part of Grice's (1975) maxim of quantity, they did not heed the second part (do not be more informative than required). In Trial C1, when they again discussed the original card set with every object unique in its basic-level category, they did not adjust their terms but continued using the same, more specific terms they had used in the B trials 52% of the time. So even though a particular object was, say, the only shoe in C1, directors referred to it with more specific terms such as *the dress shoe* more than half of the time. When a stricter, verbatim criterion for coding equivalence was applied, requiring exactly the same

word order (counting *the man's shoe* as different from *the shoe, man's*), the sequential probability of using the same terms from the last B trial to the first C trial was still 40%. So people did not follow an ahistorical, strictly informative model, but often used more informative referring expressions than necessary— that is, more informative than the ones they had used with the very same card set during the A trials. They did this even though basic-level terms are supposed to be easier to produce than subordinate terms and descriptions.

A 52% sequential probability may seem rather low, until one considers that on average, people discussed 11 other picture cards in between discussing any target twice and that to be counted as equivalent, a term had to use all of the same content words. In an investigation of a similar correspondence effect in which syntactic and lexical material from questions tends to be repeated in answers, Levelt and Kelter (1982, Experiment 5) found that there was even more correspondence between a question–answer pair when there was no intervening material than when there was intervening material. Together, these results imply that lexical entrainment is not just a local or short-term phenomenon due to priming, but that long-term memory representations are involved.

The shortening of terms over repeated referring accounts for many of the cases when pairs failed to keep using the more informative terms established during the B trials. When a pair did not use the same term in C1 as in the last B trial, but instead changed to a basic-level term in C1, the basic-level term had been contained within the more informative B term 71% of the time. Out of the total of 96 referring expressions used by the 24 pairs for the four targets during the first C trial, only 12 were neither equivalent nor shortened versions of the referring expressions used in the last B trial.

Did directors use more informative terms than they needed simply because they failed to notice that figure–ground relations had changed on the first C trial? This is unlikely because the participants witnessed the experimenter changing card sets from B to C trials. Indeed, they adjusted immediately to the change in card sets from the A to B trials, so there is every reason to suppose they noticed the card set switch on Trial C1.

Lexical entrainment. Directors were quite consistent in the terms they used in referring to the same object within the same set of objects. Pairs who matched four trials of A cards used the same referring expression in A4 as they did in A3, 81% of the time. Within four B trials, pairs were consistent in their terms as well; they used the same referring expressions in B3 as in B4 71% of the time. When they did not, the terms in the last two B trials overlapped considerably; that is, either the director proffered exactly the same phrases in both trials plus an additional phrase in one trial but not in the other, or else a phrase in one trial was slightly shortened with respect to the other, usually by virtue of omitting an adjective (as in *the big dog* and *the big red dog*). As for C trials, terms in C3 and C4 were equivalent 90% of the time. Successive references are often shortened to simpler or more efficient expressions (see, for instance, Carroll, 1980; H. H. Clark & Wilkes-Gibbs, 1986). We coded shortening whenever components of a director's reference such as morphemes, words, or modifying phrases were dropped from a subsequent reference and did not reappear again in a subsequent reference. For example, one

director used "the curved round fish with the green stripe down its back," and then "the curved round fish with the green stripe," and then "the curved round fish." At one or more points in the three B trials following B1, shortening occurred 60% of the time. So even when pairs did not use exactly the same terms for an object across trials, they almost always retained a consistent conceptual perspective. The changes they made were to simplify the conceptualization. Lengthening occurred only 8% of the time, and usually in B2, when a director sometimes offered the same term as in B1, plus an alternative term.

If frequency of use affects lexical entrainment, the participants should rely more heavily on the precedents established in the B trials, the more firmly these precedents have been established. That is, they should be more likely to continue using more specific terms like *the pennyloafer* in C1 after four B trials than after just one. And they did, 69 to 35% of the time (each analysis is reported by subjects as F_1 and by items as F_2, with MSEs rounded to two decimal places), $F_1(1, 21) = 10.80$, $p < .005, MSE = .12; F_2(1, 3) = 32.00, p < .02, MSE = .02$. The pattern was the same with the stricter, verbatim criterion, 60 to 15%, $F_1(1, 21) = 20.90, p < .001, MSE = .10; F_2(1, 3) = 150.00, p = .001, MSE = .01$. If recency alone were enough to explain lexical choice in C1, the number of B trials should not have mattered. By the same logic, the more A trials a pair had, the less likely they should be to retain terms from the last B trial on the first C trial, but this effect was not a reliable, linear trend, $F_1(1, 21) = 2.66, p < .12, MSE = .14; F_2(1, 3) = 4.20, p = .13, MSE = .05$. There was no interaction of number of A and B trials on lexical entrainment. Table 2 shows sequential probabilities of lexical entrainment from the last B trial to the first C trial, broken down by number of A and B trials.

How much did terms vary across the last three C trials—did people continue to rely on the more specific conceptualizations from the B trials, or did they revert to the basic-level term at some point in the C trials? Exactly the same terms were used across all four C trials 53% of the time. In about two-thirds of these, the terms were those that had been established during the B trials, and in about one-third, they were the basic-level terms that in most cases had been used in the A trials. Terms were shortened across the four C trials 25% of the time and lengthened only 8% of the time. If pairs had relied on simple recency, they should have kept the more specific terms throughout the C trials.

There is direct evidence that both partners contributed to

Table 2

Probabilities in Experiment 1 That a Director Used the Same Lexical Items (That Were More Specific Than the Basic-Level Term) in the First C Trial as in the Last B Trial

Number of A trials	Number of B trials		
	1	4	Overall
0	.50 (.19)	.81 (.69)	.66 (.44)
1	.25 (.19)	.69 (.63)	.47 (.41)
4	.31 (.19)	.56 (.50)	.44 (.34)
Overall	.35 (.19)	.69 (.60)	.52 (.40)

Note. Probabilities of verbatim equivalence are in parentheses.

the terms they entrained on in the B trials. When directors first proposed a term like *the pennyloafer* on those trials, their partners generally accepted the conceptualization it implied with explicit acknowledgments like *yeah* and *okay*. If they didn't accept the conceptualization for some reason, the two would work to find terms they could agree on, as here:

Director: a docksider
Matcher: a what?
Director: um
Matcher: is that a kind of dog?
Director: no, it's a kind of um leather shoe, kinda preppy pennyloafer
Matcher: okay, okay, got it

Thereafter, the director referred to this object as *the penny-loafer*. Sometimes, it was the matcher who proposed the conceptualization they agreed on:

Director: another fish, the most realistic looking one with the pink stripes, green and pink
Matcher: a rainbow trout?
Director: yeah, yeah

From then on, the director referred to the fish as *the rainbow trout*. As these examples demonstrate, partners collaborated to establish conceptualizations acceptable to both of them.

From Experiment 1, we conclude that informativeness can be overridden by historical factors and that a historical model that takes into account frequency of use as well as recency better accounts for repeated referring than one based on recency alone.

Experiment 2

References may also be shaped by provisionality. By this we mean the initial status of a term before a speaker has grounded it with an addressee: One person has proposed a conceptualization for an object, and the other has not yet agreed to that conceptualization. The provisionality of a term and its underlying conceptualization should be influenced by how certain a speaker is that her proposal will be accepted, but not by who proposed or last appealed to the conceptualization. Experiment 2 was designed to replicate Experiment 1 and to examine the provisionality of a conceptualization during interaction between speakers and addressees.

Consider the certainty of a proposal. When speakers propose a conceptualization, they cannot always be certain their partners will agree to it. If they want to refer to a pennyloafer among a set of nonshoes (Card Set A), they can be fairly confident their partners will accept its conceptualization as a shoe. But when they want to refer to the same pennyloafer in a set that includes a sneaker and a high-heeled shoe (Card Set B), they cannot be so confident their conceptualization will be acceptable. The reason is that they have alternatives. Among the alternatives used in Experiment 1 were *pennyloafer, docksider, casual shoe, brown shoe*, and *dress shoe*. The example cited earlier is a case in point. The director proposed *docksider*, but the matcher did not agree to it.

When speakers are uncertain about their proposals, they should mark them as provisional—at least if they are being cooperative. One way is with hedges—expressions such as *sort of, kind of, like*, and suffixes such as *-ish* and *-y* on adjectives. In

Experiment 1, we noticed that hedges were used just this way. Take the references to the car target item. On the A trials, there was little uncertainty about how to conceptualize it, so the directors did not have to mark their proposals as provisional. But on the B trials, there was more uncertainty, so directors sometimes marked their proposals as provisional. One director produced this series of references:

Trial B1—"a car, sort of silvery purple colored"
Trial B2—"purplish car going to the left"
Trial B3—"purplish car going left"
Trial B4—"the purplish car"
Trial C1—"the purple car"

On the first B trial, the director used two hedges, *sort of* and *-y*; on the second through the fourth, he used only one, *-ish*; and by the C trials, he was using no hedges. This dropping of hedges has been documented by Carroll (1980) for the repeated references to the fanciful drawings studied by Krauss and Weinheimer (1964) and also by H. H. Clark and Wilkes-Gibbs (1986) for repeated references to tangram figures. Experiment 2 looked at hedging as a marker of provisionality.

Speakers should be more certain about proposals that modify established conceptualizations than about those that represent new ones. One place to see such modification is in a comparison of the A and B trials. Suppose two partners agree in the A trials to conceptualize the target shoe as a shoe. When faced with three shoes in the B trials, they can either (a) modify the old conceptualization or (b) establish a new one. They should be more likely to try for a modification the more firmly the original pact had been established in the A trials. Another place we expect modification is in the C trials. If a pair entrained on terms in the B trials that contain the basic-level term, as in *dress shoe*, they should be more likely to return to *shoe* at some point during the C trials than if they had entrained on *pennyloafer*. And their use of basic-level terms in C trials should be affected by the number of A and B trials.

As for which of the two partners first proposed or last used a conceptualization, this should not matter. The adaptive factors in our model of referring presume that speakers and addressees both rely on a conceptual precedent precisely because they have established it jointly. So there should be the same pattern of lexical entrainment over repeated referring when both partners have opportunities to refer first to an object as when only one does. In Experiment 2, directors and matchers switched roles from one trial to the next. In this way both of them, as director, had ample opportunities to propose terms.

Method

Participants. Seventy-two undergraduate students (42 women, 30 men) from the State University of New York at Stony Brook volunteered to participate in exchange for research credit in a psychology class. Students who were strangers to each other were paired with members of the same sex on the basis of availability.

Materials and procedure. Experiment 2 had the same materials, task, and design as Experiment 1, except that in all sequences of four A, B, and C trials, the pairs alternated director and matcher roles. All conversations were audiotaped.

Analysis and coding. All conversation by the 36 pairs about the four target items was transcribed and double-checked for accuracy. As in Experiment 1, we computed sequential probabilities for equivalence

between references in pairs of trials of interest. Coding was done using two criteria: The first, *equivalence*, was the same criterion used throughout Experiment 1 and required that referring expressions in a pair of trials contain all the same content words, while the second criterion, *verbatim equivalence*, required that content words in referring expressions appear in the same order. So *the green curved fish* was considered the same as *the curved green fish* by the first criterion but not by the second. Both criteria ignored determiners (e.g., *a* or *the*), hedges (e.g., *kind of*), and what we call meta-references (such as the phrase *from the last time* appended to a phrase like *the big red dog*). We coded the presence of hedges and meta-references separately and also examined references for whether they contained or consisted entirely of basic-level terms. Two coders coded all of these variables for the target references of interest; their coding decisions agreed 96–100% of the time.

Results and Discussion

We considered four aspects of references to target objects: lexical entrainment, informativeness, hedges, and adaptability. Lexical entrainment and informativeness worked as in Experiment 1, whereas hedges and the adaptability of references supported the additional hypothesis that references are provisional.

Lexical entrainment. In the A trials, with only one shoe, one dog, one car, and one fish, it should have been easiest to use the unadorned basic-level terms *shoe, dog, car*, and *fish*. The base rate for using just these terms in Trial A1 was 89% (for the four targets, respectively, 79, 83, 100, and 92%). Terms in Trials A3 and A4 were identical 100% of the time (verbatim criterion), and 92% of these were basic-level terms.

In the B trials, with three shoes, three cars, three dogs, and three fish, it should have been harder to establish conceptual precedents for the targets, because there were more alternative conceptualizations possible. This variety was reflected in a greater variability of terms across pairs than within pairs. The likelihood that a director from one pair would choose the same term in the last B trial that a director from another pair had chosen in the second-to-last B trial was only 10%. This falls within the 7–18% range found by Furnas et al. (1987). Within pairs, the terms were the same 56% of the time by the equivalence criterion and 50% of the time by the verbatim criterion. Within-pair consistency between the last two B trials was higher than between-pair consistency both by the equivalence criterion, $F_1(1, 35) = 56.35, p < .001, MSE = .07; F_2(1, 3) = 31.16, p < .02, MSE = .01$, and by the verbatim criterion, $F_1(1, 35) = 35.06, p < .001, MSE = .08; F_2(1, 3) = 17.33, p < .03, MSE = .02$.

Informativeness. As in Experiment 1, matchers performed well: They made errors placing target cards only 2% of the time and corrected all but one of these later in the same trial. Directors were nearly always informative enough, adjusting immediately to the need for more specific terms in B1 when the figure–ground relationship of the target objects changed. They used more elaborate terms than simply *shoe, dog, car*, and *fish* in their first speaking turn in all but 14 of 144 B1 references. Of these 14 cases, 7 times they added information after being prompted by their partners, 3 times they gave additional installments without being prompted, and 4 times the simpler terms were sufficient because the other two objects in the same category had already been placed. When directors were

prompted by matchers, they responded with laughter, an apology, or an excuse, as here:

Director: a shoe
Matcher: uh, which shoe?
Director: ooh, forgot about that, um it's going right, it's red
Matcher: okay

As before, with a firmly established conceptual precedent, directors did not need to heed the second part of Grice's (1975) maxim, and they did not. In the first C trial, when there was again a single shoe, dog, car, and fish, directors were often overly informative, retaining the terms from their last B trial (e.g., *the pennyloafer*) 46% of the time (38% by the verbatim criterion). It was not just that individuals were perseverating on terms they had used most recently in the last B trial. Whenever there were four B trials, the two partners had switched director–matcher roles between the last B trial and the first C trial, so one partner's choice of terms retained the last conceptualization that had been used by the other partner.

By the frequency-of-use hypothesis, two partners should rely more on a conceptual precedent the more firmly it has been established, and that prediction was again confirmed. Directors were more likely to use the same term in the first C trial that had been used in the last B trial when they had had four B trials than when they had had just one, 61 to 31% by the equivalence criterion, $F_1(1, 33) = 23.35, p < .001, MSE = .07$; $F_2(1, 3) = 27.92, p < .02, MSE = .02$, and 54 to 21% by the verbatim criterion, $F_1(1, 33) = 29.61, p < .001, MSE = .07$; $F_2(1, 3) = 30.86, p < .02, MSE = .02$. The number of A trials also made a difference; directors were marginally more likely to use the last B term in the first C trial when they had had fewer A trials: linear trend, by the equivalence criterion, $F_1(1, 33) = 3.78, p = .06, MSE = .07$; $F_2(1, 3) = 8.33, p = .06, MSE = .02$, by the verbatim criterion, *ns*. Sequential probabilities of lexical entrainment from the last B trial to the first C trial are shown in Table 3.

Hedges. The provisionality of referring expressions is supported by the pattern of hedges in the A, B, and C trials. Directors should have had more reason to mark a conceptualization as provisional when a figure had alternative conceptualizations (in the B trials) than when it had a single dominant one (in the A and C trials), and they did. There was only one hedge in all 240 A trials. There were significantly more hedges in B1 than in A1, 26 versus 1%, $F_1(1, 23) = 19.26, p < .001, MSE = .02$; $F_2(1, 3) = 16.82, p < .05, MSE = .01$. In the C trials, when the two partners returned to Card Set A, they again hedged very little. The seven times people used any

Table 3
Probabilities in Experiment 2 of Using the Same Terms in the First C Trial as in the Last B Trial

Number of A trials	Number of B trials		
	1	4	Overall
0	.33 (.21)	.67 (.58)	.50 (.40)
1	.38 (.21)	.79 (.67)	.59 (.44)
4	.21 (.21)	.38 (.38)	.30 (.30)
Overall	.31 (.21)	.61 (.54)	.46 (.38)

Note. Probabilities of verbatim equivalence are in parentheses.

hedges in the 576 C trials, they had had just one B trial to match that object instead of four. Also, partners should need fewer hedges as they refine a conceptualization. That is, in B1 directors should offer provisional conceptualizations, but once partners adopt these conceptualizations, they no longer need to mark them as provisional, and the hedges should drop out. As predicted, hedges decreased from 26 to 2% of the time over the four B trials, linear trend, $F_1(1, 35) = 30.56, p < .001, MSE = .03; F_2(1, 3) = 26.44, p < .02, MSE < .01$. This pattern of hedges cannot be accounted for by a historical model based simply on recency. In a simple recency model, directors should have kept the same wording throughout the B trials, and there should be no particular reason to use or not use hedges.

In the historical, interactive model proposed so far, two partners rely on conceptual precedents even in A trials with such terms as *shoe, dog, car,* and *fish*. The more A trials a pair has had to establish precedents marked by basic-level terms, the more often they should be able to build on them in their first B trial. This leads to two predictions. First, the more A trials a pair has had, the fewer hedges they should use in their first B trial. That is, when the director has experience conceptualizing an object, she should have more confidence proposing a new conceptualization for it than if she is seeing it for the first time. This was marginally true; after zero, one, and four A trials, there were hedges in 35, 29, and 13% of their references, respectively, linear trend, $F_1(1, 35) = 6.50, p < .02, MSE = .05$; $F_2(1, 3) = 6.15, p < .09, MSE = .02$. Second, in their first B trials, directors sometimes referred explicitly to their previous conceptualization with such phrases as *the old shoe, the dog, the one from before, the car we had all along,* and *the regular old fish*. These meta-references should have decreased over four B trials, and they did, from 21 to 4%, linear trend, $F_1(1, 23) = 11.81, p < .005, MSE = .03; F_2(1, 3) = 13.17, p < .05, MSE < .01$. Also, directors should have been more likely to make meta-references after four A trials than after just one, which they did, 29% to 13%, although this difference was not quite reliable, $F_1(1, 22) = 2.38, p < .14, MSE = .07; F_2(1, 3) = 8.00, p < .07, MSE = .01$.

Adaptability. We have proposed that re-referring to an object is based on precedent, yet adaptable to the circumstances. Repeated references are often shortened as precedents are established, and people are free to modify precedents when goals or informational needs change. Recall that pairs who could firmly establish conceptual pacts over four B trials continued using the more specific B terms (such as *pennyloafer*) 61% of the time in C1 instead of reverting to basic-level terms (such as *shoe*). What about the rest of the C trials? Simple recency predicts that pairs should continue to use the terms that worked the last time. Frequency of use predicts that at least three factors should affect whether they modify their conceptual pacts during the C trials: (a) how well established any existing conceptual pact is (from the B trials), (b) whether there is an appropriate, previously established conceptual pact (from the A trials), and (c) whether the more specific B term happened to contain the basic-level term (such as *the pennyloafer shoe*).

As expected, the simple recency prediction was not supported: Pairs used increasingly more basic-level terms across C trials, linear trend, $F_1(1, 33) = 26.52, p < .001, MSE = .06; F_2(1, 3) = 48.12, p < .01, MSE = .01$. Also as expected, the

frequency-of-use predictions were supported (see Table 4). Pairs were more likely to revert to basic-level terms in C trials when there had been only one B trial rather than four, $F_1(1, 33) = 32.93, p < .001, MSE = .23; F_2(1, 3) = 67.05, p < .005, MSE = .04$. The more A trials there had been, the more often they used unadorned basic-level terms during the four C trials, linear trend, $F_1(1, 33) = 5.76, p < .05, MSE = .06; F_2(1, 3) = 144.23, p < .001, MSE < .01$. According to our historical model, during the C trials these pairs could re-evoke the conceptual precedents they had firmly established during the A trials. Finally, pairs reverted to the basic-level term at some point during the C trials 49% of the time when their B terms had contained the basic-level term (as with *the big dog*) and 14% of the time when they had not (as with *the Irish setter*), $F_1(1, 23) = 27.20, p < .001, MSE = .16; F_2(1, 3) = 58.32, p = .005, MSE = .03$.[3] In these cases, reverting to the basic-level term may have been due to simplifying an existing conceptualization, rather than establishing a new one.

Even though partners kept the same director–matcher roles in Experiment 1 and switched roles in Experiment 2, the patterns of results remained the same: Terms in the last B trial were equivalent to those in C1 52% and 46% of the time, respectively, and having more B trials made this even more likely, 69% and 61% of the time, respectively. In Experiment 2, the number of A trials had a reliable effect as well, although this effect was not quite reliable in Experiment 1 (Experiment 2 had 50% more pairs of participants than Experiment 1). So lexical entrainment in this task did not depend on which partner had originally proposed a conceptualization.

Experiment 3

In Experiment 3 we examined the idea that conceptual pacts are established with particular individuals. Pairs of people discussed sets of picture cards for eight trials, as in the B and C trials of Experiments 1 and 2. In the first four (B) trials, the targets were not unique in their basic-level categories (there were three shoes, three dogs, three cars, and three fish), but in the last four (C) trials, they were. Pairs participated in one of two conditions. In the same-partner condition, directors matched cards with the same matcher in both the B and the C trials. In the switch-partner condition, they received a new matcher for the C trials. If directors form conceptual pacts with specific matchers, they should use the same terms in Trial

Table 4
Effect of Frequency of Use: Probability of Reverting to Unadorned Basic-Level Terms in C Trials as a Function of Number of A and B Trials in Experiment 2

Type and number of trials	Trial C1	Trial C2	Trial C3	Trial C4
A trials				
0	.17	.31	.38	.46
1	.31	.42	.60	.56
4	.56	.58	.58	.63
B trials				
1	.47	.64	.68	.71
4	.22	.24	.36	.39
Overall	.35	.44	.52	.55

C1 as in Trial B4 more often when continuing with the same partner than when faced with a new partner. And in the C trials they should be more likely to rely on precedents established in the B trials and less likely to revert to basic level terms like *shoe* and *fish*.

Partner-specific effects have two possible origins. First, speakers may tag a conceptual pact such as *loafer* as shared with a particular partner (see, e.g., H. H. Clark & Marshall, 1978). So when directors meet the same matcher again on the C trials, the identity of the matcher serves as an additional memory cue for the old conceptualization. When directors meet a new matcher on the C trials, that cue is missing, so they should be more likely to abandon the precedent and offer basic-level terms. Or second, matchers who have not yet formed a conceptual pact with directors should expect the basic-level term *shoe* on the first C trial. And if directors do not use *shoe*, these matchers should encourage them to do so over the rest of the C trials. A conceptual pact, then, need not be represented explicitly but may emerge from the conceptual coordination of two people interacting.

Method

Participants. Fifty students from the State University of New York at Stony Brook (39 women and 11 men) volunteered to participate to fulfill a course requirement or for a $5 honorarium. Students were grouped on the basis of availability and participated in groups of either 2 or 3, to form 10 same-partner groups and 10 switch-partner groups. Second matchers waited in a separate room during the experiment until they were needed.

Materials. Experiment 3 used the same picture cards and a subset of the tangram distractors from Experiments 1 and 2. As before, the same four targets (a shoe, a dog, a car, and a fish) appeared in the unique card set as lone exemplars of their basic-level categories and in the nonunique card set with two other category members. This was done by combining Sets 1A and 2A from Experiments 1 and 2 to make one set of 20 cards that contained all of the same (unique) picture cards and eight of the tangram distractors, and Sets 1B and 2B to make another set of 20 cards with all of the same (nonunique) picture cards and eight of the tangram distractors.[4] We continue to refer to the nonunique trials as B trials (even though in Experiment 3 there were not any A trials) and to the unique trials that followed them as C trials.

Procedure. A director and a matcher did four trials matching the nonunique card set. Then there was a short pause during which the experimenter changed the card sets; in the switch-partner condition, the first matcher was taken out of the room and a second matcher was brought in and instructed in the task. Then the unique cards were arranged in the director's and matcher's frames. The pause in between the last nonunique trial and the first unique trial was roughly the same for both conditions. Before starting the first trial with a particular card set, the director was instructed to look over the new card set while the matcher's cards were being arranged, and she had at least 30 s to do so.

Analysis and coding. All speech about the four target items was transcribed and checked. The trials of particular interest were the last

[3] The *df*s for F_1 result from the fact that for the four target items, 24 of the 36 pairs entrained on some B terms that included the basic level and some that did not.

[4] We combined the sets to be able to make only one partner switch in the switch-partner condition during the experimental session. Recall that in Experiments 1 and 2, pairs went through the A, B, C (unique, nonunique, unique) sequence twice, each sequence with its own distinct set of target and distractor cards.

five trials, B4 and C1–C4. Two coders rated references to target items for equivalence across trials B4 and C1, where equivalence meant that two references included all the same content words (so *a shoe, man's* was considered equivalent to *a man's shoe*). They also coded all references to target items in the four C trials as either unadorned basic-level terms or as more informative. For the variables coded, the two coders' decisions were the same 98–100% of the time.

Results and Discussion.

Lexical entrainment. As predicted for conceptual pacts, directors used the same terms in Trial C1 as in Trial B4 more often when continuing with the same matchers than after switching matchers. The difference was 48 to 18%, reliable by subjects and marginally so by items, $F_1(1, 18) = 5.18, p < .05$, $MSE = .09$; $F_2(1, 3) = 6.75, p = .08$, $MSE = .03$. Yet when switching matchers, directors did not always abandon the conceptualizations they had formed in the B trials and start over; sometimes they proposed the conceptualizations they had arrived at with the former matcher in the B trials, but in an elaborated form. When we considered only the C1 references that were neither basic-level terms nor equivalent to B4 references, we found that directors with new matchers lengthened their references 67% of the time, whereas directors with old matchers lengthened their references only 22% of the time, $F_1(1, 15) = 6.20, p < .05, MSE = .14$; $F_2(1, 3) = 54.00, p = .005, MSE = .01$, shortening them the rest of the time.

Basic-level terms. Although there was no reliable difference in the total number of basic-level terms used in C trials for the same-partner and switch-partner conditions, there was, as expected, an increase in references consisting of unadorned basic-level terms over the four C trials, linear trend, $F_1(1, 18) = 13.96, p < .002, MSE = .04$; $F_2(1, 3) = 71.04, p < .005$, $MSE < .01$. This increase was small in the same-partner condition, 15 to 23%, but dramatic in the switch-partner condition, 20 to 55%, linear trend of interaction, $F_1(1, 18) = 5.49, p < .05, MSE = .04$; $F_2(1, 3) = 14.93, p < .05, MSE = .01$. Speakers with the same partners were more likely to continue to rely on conceptual pacts on the C trials even when this meant using overinformative terms. That is, they continued to use *loafer* when they could have changed to *shoe*. Speakers with new partners abandoned their former conceptualizations for simpler ones. That is, over the four C trials, they moved from terms like *loafer* to *shoe*. Table 5 shows the percentages of unadorned basic-level terms over the four C trials.

The references in our task emerged not from solitary choices by the director, but from an interactive process by both director and matcher. Evidence for this is found in the form and timing of both partners' utterances. Consider one director's Trial C1 with a brand-new matcher:

> Director: nine is a fish [1.5 s pause]
> kinda green and pink
> Matcher: done

Table 5
Probability of Reverting to Unadorned Basic-Level Terms in C Trials as a Function of Partner in Experiment 3

Partner	Trial C1	Trial C2	Trial C3	Trial C4
Same	.15	.18	.23	.23
Switch	.20	.27	.45	.55

The director first offered the unadorned basic-level term *fish*. If the matcher had acknowledged it immediately, the director might have left it that way. But because the matcher paused so long, the director added a phrase, making the reference more informative than the basic-level term would have been. In contrast, here is another director's Trial C1 with a new matcher (where pairs of asterisks denote overlapping speech):

> Director: number 11 is a pair—it's um, sorry, sorry, sorry, it's a fish with *different colors*
> Matcher: *yeah* okay
> [Thereafter, the director used *the fish*]

Here again the director offered the basic-level term *fish*, but as she began to add more information, the matcher interrupted with *yeah* to indicate he did not need it; the term *fish* was informative enough. Or consider another director's Trial C1 with a new matcher:

> Director: number 3 is a, a car, it's facing the—
> oh yeh, yeh, it's a car
> did you get that?
> Matcher: yes

Although this director began to give more information than the basic-level term *car*, he corrected himself, showing he was fully aware of his new matcher's informational needs.

In summary, Experiment 3 shows a partner-specific effect on repeated referring. Speakers with continuing addressees appeal to conceptual pacts they have already established, even when these are overinformative. Speakers faced with new addressees sometimes begin by proposing the conceptualizations they have established with previous addressees, but they rapidly accommodate to their new addressees.

General Discussion

These experiments support a historical model of referring over an ahistorical model. When people refer to an object in conversation, we have proposed, they establish a conceptual pact, a temporary agreement about how they and their addressees are to conceptualize that object. It may be to conceptualize a particular shoe as a *shoe* or *pennyloafer* or *casual shoe* or *our old shoe* (i.e., the shoe we talked about before). Once they establish a conceptual pact, either partner can appeal to it in referring to the shoe a second, third, or fourth time. One consequence is lexical entrainment, the repeated use of the same or closely related terms in referring to an object on successive occasions.

Conceptual pacts have several characteristics documented in these experiments.

1. In conversation, conceptual pacts are established by speakers and addressees jointly. Ordinarily, speakers propose a term for a conceptualization, such as *docksider* or *pennyloafer*, and their partners agree to it, ask for an alternative, or suggest a substitute until the two of them arrive at a jointly agreeable conceptualization (see also H. H. Clark & Schaefer, 1989; H. H. Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992).

2. When speakers propose a conceptualization, they often mark it for how confident they are that it will be understood and adopted by their addressees. When they are confident, they simply presuppose a conceptualization, as with *the shoe*.

When they are not so confident, they may mark a reference as provisional by using hedges, as in *a kind of, um, leather shoe, kinda preppy pennyloafer.* Our directors were more likely to hedge when there were many alternative conceptualizations (during B trials) than when there was a single dominant one (during A and C trials).

3. People do not establish conceptual pacts all at once, but often little by little. We have assumed that the more times two people refer to an object, the stronger any resulting conceptual pact is likely to be. So the more firmly two partners establish a pact, the more likely they are to appeal to it later, and to appeal to it with confidence. Frequency of use better explains our data than does simple recency. In Experiment 2, the more A trials our partners had, the more likely they were to appeal to the A conceptualizations in the B and C trials and the less likely they were to hedge during B trials. As they proceeded through the B trials, they used fewer hedges. Likewise, the more B trials they had had, the more likely they were to appeal to the B conceptualizations in the C trials.

4. Conceptual pacts are accessible to both speakers and addressees. The directors in our study appealed to conceptual pacts not only when they themselves had proposed the terms marking them (Experiment 1) but also when their partners had (Experiment 2).

5. Speakers form conceptual pacts with particular address-ees. Addressee effects have been found elsewhere, for side participants versus addressees (Wilkes-Gibbs & Clark, 1992) and for overhearers versus addressees (Schober & Clark, 1989). Both these role-specific effects and the partner-specific effects in Experiment 3 seem to have emerged as a result of the feedback that addressees provided during the interactive grounding process.

Carroll (1980) has proposed an account of referring in which, like our account, the participants in a conversation share responsibility for naming things. In Carroll's account, speakers propose references and, once their partners ratify them, use them consistently thereafter. He proposes that names emerge abruptly during repeated references that usu-ally start out as longer descriptions: "The final name pops out as a response by one or the other participant, is then accepted by the partner, and is used thereafter" (Carroll, 1980, p. 315). Our experiments show, however, that references are more adaptable than this suggests. The adaptability is of two sorts. Speakers could modify their references gradually over re-peated trials with a card set. Their references often became simplified and more efficient over time and were not abruptly replaced by names. On the other hand, speakers could modify their references abruptly as they moved from the last A trial to the first B trial—in response to changing goals and informa-tional needs. Sometimes, in the C trials, they reverted to the conceptual pact they had established earlier, in the A trials. Both gradual and abrupt changes are consistent with concep-tual pacts, which are provisional. This adaptability is not accounted for by the establishment of conventional names, as in Carroll's proposal.

Finally, lexical choice in a communicative setting can display not only agreement, but also disagreement. Our experiments used objects that were as simple as possible in order to bring this phenomenon under laboratory control; even in such simple situations, people do not always reach shared conceptu-alizations. The Appendix shows an example from Experiment 2 when a director and matcher failed to lexically entrain over seven trials: As they changed roles their references showed a pattern of two distinct, alternating terms, appearing to index two distinct conceptualizations. Examples like this show the sense in which the usual agreements about conceptualizations are genuine pacts. The phenomenon is even more striking when there is more at stake. In the 1975 trial of a Boston physician who was charged with murder for performing an abortion, the prosecutor referred to the aborted fetus through-out as a baby, whereas the defense lawyer referred to it throughout as a fetus (Danet, 1980). Each lawyer made a point of not accepting the conceptualization proposed by the other, and presumably this point was not lost on the jury.

What we have shown, then, is that lexical entrainment in conversation is better accounted for by conceptual pacts than by informativeness or by recency alone. In our proposal, people establish, track, and update conceptual pacts (i.e., provisional agreements about how they are to conceptualize things). When speakers refer, we claim, they are trying to do more than get their addressees to pick out the right referent. They are trying to get them to pick out the right referent under the right conceptualization, for what they have to say about the referent is specific to that conceptualization.

References

Bolinger, D. (1977). *Meaning and form.* London: Longman.

Brennan, S. E. (1990). *Seeking and providing evidence for mutual understanding.* Unpublished doctoral dissertation, Stanford Univer-sity.

Brennan, S. E., Ries, P. S., Rubman, C., & Lee, G. R. (1996). *The vocabulary problem in speech and language interfaces.* Manuscript submitted for publication.

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

Brown, R. (1958b). *Words and things.* Glencoe, IL: The Free Press.

Carroll, J. M. (1980). Naming and describing in social communication. *Language and Speech, 23,* 309–322.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Erlbaum.

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

Clark, H. H., & Clark, E. V. (1979). When nouns surface as verbs. *Language, 55,* 430–477.

Clark, H. H., & Marshall, C. (1978). Reference diaries. In D. L. Waltz (Ed.), *Theoretical issues in natural language processing, TINLAP-2* (pp. 57–63). New York: Association for Computing Machinery.

Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes, 2,* 19–41.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13,* 259–294.

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

Cruse, D. A. (1977). The pragmatics of lexical specificity. *Journal of Linguistics, 13,* 153–164.

Danet, B. (1980). 'Baby' or 'fetus'?: Language and the construction of reality in a manslaughter trial. *Semiotica, 32,* 187–219.

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

Ford, W., & Olson, D. (1975). The elaboration of the noun phrase in children's descriptions of objects. *Journal of Experimental Child Psychology, 19,* 371–382.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communications. *Communications of the Association for Computing Machinery, 30,* 964–971.

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

Garrod, S., & Clark, A. (1993). The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes, 8,* 101–126.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition, 53,* 181–215.

Grice, H. P. (1975). Logic and conversation (from the William James lectures, Harvard University, 1967). In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form and use in context: Linguistics applications.* Washington, DC: Georgetown University Press.

Krauss, R. M. (1987). The role of the listener: Addressee influences on message formulation. *Journal of Language and Social Psychology, 6,* 81–98.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science, 1,* 113–114.

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

Kraut, R. E., Lewis, S. H., & Swezey, L. W. (1982). Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology, 43,* 718–731.

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

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology, 14,* 78–106.

Levinson, S. (1983). *Pragmatics.* Cambridge, England: Cambridge University Press.

Mangold, R., & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology, 7,* 181–191.

Murphy, G. L., & Smith, E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior, 21,* 1–20.

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

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics, 27,* 89–110.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

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

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

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language, 31,* 183–194.

Appendix

In this (atypical) example, Partners X and Y fail to lexically entrain on a specific term for the target object *dog;* in fact, their terms appear to mark two distinct perspectives. In Trial B1, X proposes the term *lab,* and Y asks for clarification. When Y is next the director, she counterproposes *red dog.* In the next five trials, they each stick to their own terms, while understanding each other's term perfectly well. On Trial C4, Y finally adopts X's term.

Trial	Speaker	Reference to target item
B1	X	next one is a lab, red
	Y	a what?
	X	a red lab
	Y	uh huh
B2	Y	and then the red dog
	X	[no response]
B3	X	next one is the red lab
	Y	uh huh
B4	Y	then the red dog
	X	[no response]
C1	X	red lab
	Y	uh huh
C2	Y	then the red dog
	X	[no response]
C3	X	red lab
	Y	uh huh
C4	Y	then the red lab
	X	[no response]