ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Memory and Language 49 (2003) 201–213

Journal of Memory and Language

www.elsevier.com/locate/jml

When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions

Charles Metzing and Susan E. Brennan

State University of New York at Stony Brook, Stony Brook, NY 11794-2500, USA

Received 12 March 2002; revision received 18 February 2003

Abstract

When two people in conversation refer repeatedly to objects, they typically converge on the same (or similar) referring expressions. The repeated use of expressions by people in the same conversation has been called *lexical entrainment*. Lexical entrainment may emerge from the precedent of associating objects with expressions (and the perspectives they encode), or else from achieving *conceptual pacts*, or temporary, flexible agreements to view an object in a particular way (in which case the precedent is encoded as specific to a particular partner). We had people interact with a confederate speaker, entraining on shared perspectives (e.g., ''the shiny cylinder'') during repeated references to objects. Then either the original speaker or a new speaker used either the original expression or a new one (''the silver pipe'') to refer to the previously discussed object. Upon hearing the original expressions, addressees looked at and then touched the target objects equally quickly regardless of speaker. However, with new expressions, there was partner-specific interference: addressees were slower to look at the object when the new expression was uttered by the original speaker than when the new expression was uttered by the new speaker. This suggests that the representations in memory from which entrainment emerges do encode a partner-specific cue, leading addressees to expect that a speaker should continue to use an entrained-upon expression unless a contrast in meaning is implicated. There appears to be no such interference when a new partner uses a new expression.
© 2003 Elsevier Science (USA). All rights reserved.

When speakers refer to something, be it an everyday object or an exotic, unfamiliar one, they are faced with many choices. One person's *valuable antique* may be another's *yard sale fodder* Consider the following exchange:

Mike: Tina, I have something to tell you.
Tina: mmm?
Mike: uh, while you were out before—I was dusting—and I guess I broke one of your figurines.
Tina: oh no! which one?
Mike: it was the one—sort of blue, y'know, like a girl dancing.
Tina: no! well at least it wasn't the one my grandmother gave me.

Successful referring requires that the speaker and addressee be able to take, at least for the moment, the same perspective on a referent. In this exchange, Mike and Tina come to the belief that they are both talking about the same object. The expressions they choose reflect, at least in part, how they conceptualize the object at that moment, as well as how they propose to distinguish it from other potential referents (Brown, 1958a, 1958b; Olson, 1970). Once a referring expression has been presented by a speaker, it may be accepted and taken up by the addressee, or it may be adjusted, depending on whether the addressee understands and accepts the perspective it expresses. Once both partners have enough evidence to believe they are talking about the same thing, the mapping between the referent and the perspective has been *grounded*. Unless the context changes, they tend to use the same referring expression

E-mail addresses: cmetzing@ic.sunysb.edu (C. Metzing), susan.brennan@sunysb.edu (S.E. Brennan).

0749-596X/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0749-596X(03)00028-7

again or else a similar but shortened one when they continue to talk about the same referent (Brennan & Clark, 1996; Carroll, 1980). This process of converging has been demonstrated repeatedly in referential communication studies and is known as *lexical entrainment* (Brennan & Clark, 1996; Garrod & Anderson, 1987). As a result, the variability of referring expressions *across* conversations is much higher than that *within* conversations (Brennan & Clark, 1996). Despite this consistency, the mental representations that underlie referring must be flexible. After all, most referents can be mapped to a variety of labels, the figure-ground relationships among potential referents change frequently, and perspectives are modified to meet the changing informational and pragmatic needs of a situation.

It is widely assumed that language users follow Grice's Cooperative Principle and its maxims: Quality (be truthful), Quantity (say as much as is needed; do not say more than is needed), Manner (be clear, brief, and orderly), and Relation (be relevant) (Grice, 1975). The predictions from these maxims regarding choice of referring expressions are not always obvious. Studies of referential communication have found that speakers try to be at least as informative as they need to be. However, after they have set precedents with their conversational partners, speakers often say more than is needed to uniquely identify an object (Brennan & Clark, 1996). For instance, pairs of speakers who had repeatedly used terms like *pennyloafer* to distinguish one particular shoe from others tended to continue using the entrained-upon term even when the background set changed so that it no longer contained any other shoes (and so the basic level term *shoe* would have been sufficient to uniquely identify the object). This tendency was stronger, the stronger the precedent (that is, the more times the pair had referred to the object previously using the more informative term; Brennan & Clark, 1996). Although this finding is not what would be predicted according to Grice's Maxim of Quantity, it *is* consistent with the Maxim of Manner (Grice, 1975), by which speakers ordinarily strive to be brief, to be clear, and to avoid ambiguity. According to the Maxim of Manner, speakers should not abandon a perspective without good reason; therefore addressees should expect them to continue to rely on previously established perspectives, using the same referring expressions (though perhaps in an abbreviated form) on the next occasion of referring.

Entrainment is even more strongly predicted from another well-known pragmatic principle, E. Clark's *Principle of Contrast* (1987). When a speaker who has previously relied on a particular referring expression suddenly uses a brand new expression that seems to invoke a different conceptualization, a contrast in meaning is implicated. The addressee would not reach the same implicature upon hearing the new expression produced by a different speaker. Consistent with this principle, Brennan and H. Clark (1996) proposed that a precedent established in referring may be partner-specific, representing a *conceptual pact* to view the referent in a particular way. In their Experiment 3, when speakers had established a precedent with one partner and then met a *new* partner with whom they continued to discuss the same referent, they were more likely to modify their previous referring expression to be only as specific as it needed to be (e.g., switching to the unadorned basic level term *shoe*) than if they had continued to discuss the object with the original partner (in which case, they tended to keep using the over-informative *pennyloafer*). Certainly this apparent partner-specificity in referring could unfold gradually due to the inferences that speakers make about the acceptability of an expression to their partners. It may also occur if the original addressees were more accepting of over-informative expressions than new addressees, thus shaping speakers' behavior via feedback (as suggested by Brennan & Clark, 1996). What that program of research did not establish was whether what appear to be partner-specific choices can emerge, at least in part, from underlying representations in memory in the earliest moments of processing.

The issue of partner-specificity was pursued by Barr and Keysar (2002), who argued that the precedents from which entrainment emerges are independent from particular partners in conversation. Their experiments were conducted on addressees who first entrained on perspectives with a (confederate) speaker who produced scripted referring expressions. In their Experiment 2, once the speaker and addressees had established precedents for referring expressions, the addressees heard the same expressions again, either from the original (live) speaker or else from a new speaker (a prerecorded voice delivered through an earphone). Barr and Keysar reasoned that if entrainment is indeed based on a partner-specific representation, the precedent established with the original speaker should be inhibited when a familiar referring expression is delivered by a new speaker, causing addressees to be slower to gaze at and reach for the target object. However, their addressees were equally fast to gaze at and reach for target objects, regardless of who the speaker was. The conclusion from this experiment was that speakers and addressees rely on precedents for referring because they are available in memory, and that these precedents are not represented in any partner-specific manner.

Here we argue that the latter part of this conclusion was premature. First, from the standpoint of conversational pragmatics, an addressee's processing should not be inhibited when a new speaker happens to use the same expression as a previous speaker. To revisit our hypothetical example, imagine that a new character, Tina's other friend Harry, drops in later that day and asks, "What's the deal with the dancing girl?" That Harry happens to take the same perspective Mike did,

even though he was not present at the initial episode, should not be particularly surprising to Tina, nor should she be hindered from determining that he is referring to the same object that Mike was earlier. After all, she most likely has the episode well-encoded in memory, and so the referring expression will provide a good cue. Moreover, Harry might have chosen the same expression by chance, or he might have heard about the episode from Mike. On the other hand, it should not be particularly surprising if Harry were to ask "What's the deal with that broken tchotchke?" as long as this new expression proposes a reasonably accessible perspective that can be easily mapped onto the object.

The interesting question for partner-specificity is: What happens when a precedent is abandoned by a particular speaker? Consider the impact on Tina's processing if *Mike* were to return, announcing, "Guess what—that broken tchotchke can be fixed!" Of course, such an utterance is possible and felicitous on most theories of language use, and Tina would probably be able to figure it out. But Mike would be abandoning the precedent they had established earlier, and by the Principle of Contrast, should have a reason for doing so. If the *dancing girl* precedent set with Mike had been encoded by Tina in a partner-specific manner, then the new referring expression *broken tchotchke* should be interpreted more slowly if Mike uttered it than if Harry did. Alternatively, if the use of precedent in referring is truly partner-independent, then who the speaker is should not matter at all when a new referring expression encoding an entirely new perspective is used in the absence of any discernible reason for doing so.

Barr and Keysar's Experiment 2 omitted this crucial comparison. Their Experiment 3 tested listeners' comprehension of new expressions (e.g., *car*) spoken by original or new speakers after the original speaker had entrained with the listeners on a different term for that object (*sportscar*).[1] Their critical test involved listeners selecting from only two pictures—the target object (the sportscar) or another equally familiar item whose entrained-upon name began with the same syllable as the basic level term for the distractor (*carnation*). Upon hearing the new term *car*..., listeners gazed to an equal degree at the carnation, and this occurred regardless of who was speaking (whereas a partner-specific effect would have shown more competition from *carnation* with the original speaker than with the new speaker). However, we argue that this failure to find a difference does not provide convincing evidence against partner specificity in referring. If, as we propose, entrainment is supported by an underlying episodic representation that associates a referent, a referring expression (and the perspective it encodes), and other relevant information about the context of use (such as who a partner is), then a potential partner-specific memory cue might very well be swamped by the strong and highly coincidental cue of a competitor's label belonging to the same auditory lexical cohort. Demonstrating that one memory cue is not strong enough to overcome interference from another cue does not rule out the possibility that the weaker cue still exists.

Here, we are interested in the situation in which a speaker unexpectedly breaks a conceptual pact by using a new expression that encodes a distinctly different perspective at a comparable level of detail. The experiment we report tests the prediction that entrainment representations encode partner-specific information. We had addressees follow instructions spoken by a confederate speaker who referred in a pre-scripted manner to the same objects during each of three rounds of a matching task, so that the pairs entrained on referring expressions for the target objects. After three rounds with a set of objects, the speaker then left the room briefly and returned, or else a new confederate speaker entered, for a fourth round. In this round, the (original or new) speaker used either the original expression or an entirely new one. Addressees' eyegaze over the set of objects and reaching behavior were monitored in order to measure the time course by which they mapped the critical expression onto the target referent.

Method

Pairs of people, one a nave participant and the other an experimental confederate, did a matching task together. The naïve participant served as the Matcher and the confederate served as the Director. The task involved the Director giving the Matcher instructions to move small objects to new locations within a vertical 5×5 grid of open cubbyholes, in order to match a picture only the Director could see (see Fig. 1). There were eight different sets of objects, and each set contained one target object. For each of these sets, pairs

[1] Note that the changes from subordinate- to basic-level terms in at least some of Barr and Keysar's Experiment 3 stimuli would not actually represent *breaking* a precedent. For a term to evolve from *sportscar* to *car* within a conversation can result from a natural progression toward efficiency rather than a break with precedent. Even after speakers entrain on terms, they often continue to shorten them (indeed, Brennan & Clark, 1996 found that when speakers had entrained on subordinate level terms that included the basic level term, such as *the sportscar*, they were significantly more likely to revert to the basic level term *car* than were those who entrained on terms such as *the hotrod*). Another aspect of Barr and Keysar's paradigm that could be problematic is that the critical trials of listeners' comprehension of *car* also involve figure-ground changes (the referent is being distinguished from a different background set than when it was discussed earlier), and so listeners may have different expectations about the entrained-upon term.

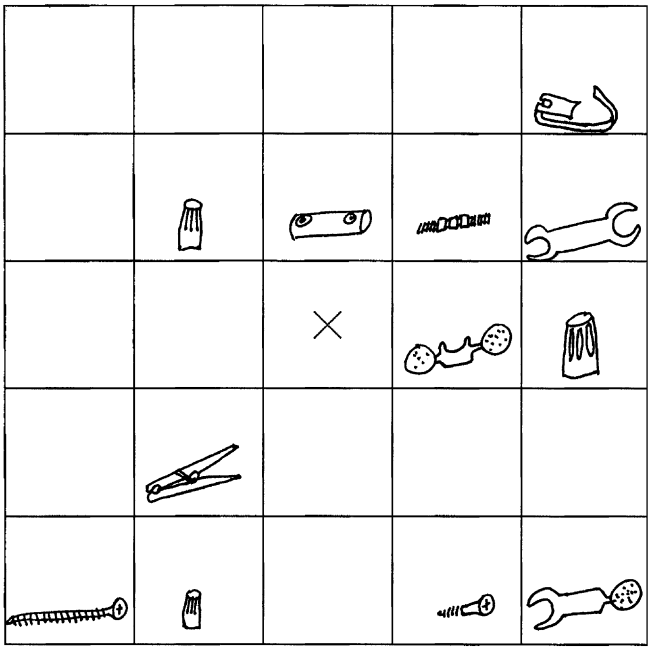

Fig. 1. Schematic of display and objects.

did a total of four consecutive rounds, within each of which the Matcher moved 6–7 of the 12 objects, including the target one. This ensured that the pair would have enough occasions of referring (during the first three rounds) to establish a conceptual precedent for each object and to entrain on an appropriate referring expression. In each round, the Director referred to each object at least once, either as an object to be moved or as a landmark relative to another object to be moved. After each third round, the Director left the room and then either re-entered or else a new Director entered for the fourth round. Each fourth round contained one critical instruction in which the Director used either the original or a new referring expression for the critical object.

Design

Each Matcher participated in eight critical (fourth) rounds, one per set of objects: in two critical rounds, the original Director continued to use the original entrained-upon expression, in two the original Director used a new expression, in two the new Director used the original expression, and in two the new Director used a new expression. All participants matched the sets of objects in the same order. The conditions for the critical rounds were counterbalanced to four versions so that each critical round appeared in all four cells of the design, and each participant experienced one of these versions.

Participants

Seven undergraduates (3 males and 4 females) from the State University of New York at Stony Brook participated in a norming task, and 24 different undergraduates (11 males and 13 females) from the same population of students participated as matchers in a referential communication study. All were native speakers of English and received research credit in a psychology course. The role of the original Director was played by the same male for all experimental sessions, and that of the new Director was played by the same female.

Materials

An open cardboard grid of 5×5 cubbyhole-like boxes, each box approximately 3 in. square and 2.5 in. deep, was positioned in between the Director and Matcher so that both could see each other as well as the objects. The center box was blocked with paper and marked with a cross to provide a fixation point for the Matcher. There were eight completely different sets of 12 objects consisting of small toys and pieces of hardware, one set for each of the critical rounds. For critical rounds, target objects were located in one of the eight boxes closest to the central fixation cross; therefore no other object was ever located in between the central fixation cross and the target, nor was another similar object ever located along the same radius

outward from the central fixation cross. Each object was small enough to fit easily into a cubbyhole box. Each set was kept out of the Matcher's view except for when it was needed.

During the referential communication task, the Director referred to objects using referring expressions in which the content words were standardized according to a script. The objects were chosen so that they could not be distinguished from the rest of the set simply by being labeled with a basic level term, but would need at least one adjective and/or a subordinate term to be uniquely identified (e.g., *shiny cylinder*, a.k.a. *silver pipe*). The critical referring expressions were first normed. Seven people were asked to name each object in the context of the other objects in its set. An adjective and a noun were selected from these names to construct the scripted expressions. Both words in each scripted expression had been produced from 1–7 times during norming (not always in the same phrase). For the critical rounds, the new and original expressions were chosen such that, on average across each cell of the design, neither new nor critical expressions were dominant in mean frequency over the other (in 2 cases the words in the original expression were mentioned more times than those in the new expression, in 3 cases the words in the new expression were mentioned more often, and in the remaining 3, the words in both expressions were mentioned equally often during norming). See Appendix for a list of the new and original expressions.

We used relatively large sets of objects rather than systematically including just one or a few highly similar competitor objects (such as Barr and Keysar's Experiment 3, which used a single competitor per set). Nevertheless, six of our eight sets contained one additional object onto which the adjective in the new target expression *could* have easily been mapped; for instance, for the object that was first called *the shiny cylinder* and later, *the silver pipe*, there existed in the background set a hook that was also silver, but that was actually referred to as *the big hook* (the other two sets contained no single good potential competitor). However, the adjectives in the new and original expressions for a given target object were never used for any other object in its set of 12. Within the sets, all adjective-noun combinations were logically sufficient to identify the objects that they were used to refer to.

Each critical utterance was scripted to be produced in two parts; the first part identified the object to be moved (e.g., *next is the shiny cylinder*), and the second part specified the target location (e.g., *put it below the Lego bridge*). This was both natural, in terms of the installments speakers produce when referring to objects in the visual environment, and necessary, to ensure that our measurements were restricted to only those looks used to identify the object as opposed to those for repositioning it. The Director never revealed the target location until the Matcher had identified and touched the object to be moved.

Even though the norming made it likely that the *content* of the expressions presented by the confederate Director would seem plausible to the naïve Matchers, there was still the problem of scripting the *process* of referring so that it would seem as natural as possible. We addressed this issue in two ways. First, we based the script on a corpus of spontaneous conversations in which both directors and matchers were naïve participants (from Brennan & Clark, 1996). In spontaneous speaking, both parties ordinarily have the opportunity to contribute to and ratify or modify the perspective that one of them has proposed, and they mark having achieved a shared perspective by re-using the same or very similar referring expression (often slightly shortened) when they re-refer to the object. When speakers first propose a referring expression for an object for which there seem to be several likely perspectives, they often hedge, as with the italicized expressions in the utterance: "Alright do you see the *kind of* gray toy plastic connector *thing* on the top row?" As two people entrain on a shared perspective, the hedges drop out (Brennan & Clark, 1996). In all, hedges were used in 81% of the referring expressions in Round 1 when referring expressions were first presented by the Director, and they were omitted from Rounds 2 and 3.

The second way in which we addressed the issue of naturalness was to ensure that the Matchers felt that they had ample opportunity to contribute to the referring process. After the Director issued the first installment of the instruction (e.g., *Now take the shiny cylinder),* which constituted the period during which we were interested in measuring the Matcher's eye movements, the Director responded naturally to the Matcher's action (e.g., by saying *yeah, that one* after the Matcher had picked up the object). Then the Director continued by specifying the target location. If a Matcher asked for clarification or paused, expressing confusion, the Director responded spontaneously and returned to the script after the Matcher appeared to be satisfied. Thus, the critical referring expressions presented by the Director were precisely scripted, and the rest of the utterances were relatively spontaneous. Table 1 contains an example of the four rounds of conversation for one of the critical items.

To assess just how natural and interactive these partially scripted sessions were, we counted the naïve Matchers' responses for each of the Director's instructions during Round 1 (this was the round during which Matchers did the most speaking). In response to the Director's instructions, Matchers simply picked up the target object 78% of the time. The rest of the time, touching the object was preceded by the following behaviors: 18 hesitations (either silent or accompanied by a sound such as "um..."), 51 clarification questions, and

Table 1
Transcript of the discussion of a single target object by a single matcher and director in rounds 1–4[a]

Round 1	
D1:	then.. see that car that has like...blueprints painted on the side of it sorta...it has like
M:	this? *(pointing out another car in the array of objects)*
D1:	no..it's that other car it has like something printed on the side
M:	*(grasps target but does not remove it from its position)*
D1:	yeah that one...
M:	*(removes target from position)*
D1:	umm.. put that in the upper corner next to the car with the number fifteen
M:	*(places car in new position)*
D1:	yeah...right
Round 2	
D1:	a:nd the last thing...take the blueprint car
M:	*(picks up target)*
D1:	a:nd and put it.. between the clothespin and that silver muffler thing
M:	*(places target in new position)*
D1:	good
Round 3	
D1:	next take the blueprint car
M:	*(picks up target)*
D1:	a:nd move it up between the car number seven and the cowboy
M:	*(places target in new position)*
D1:	okay
Round 4	
D2:	a:nd now take the blueprint car
M:	*(picks up target)*
D2:	and put it above the cowboy
M:	*(places target in new position)*
D2:	okay

[a] *Note.* A question mark marks rising intonation of the sort used in questions, a colon marks lengthening of part of a word, and an elipsis (...) marks pausing. Material in parentheses describes the Matcher's actions, which were visible to the Director.

184 gestures that consisted of silently pointing to or touching an object, or else gesturing accompanied by a clarification question such as *this?* These types and rates of responses are comparable to those in other studies in which both partners are naïve (and interaction is entirely unscripted), where they have some visual co-presence, and where the objects being matched are somewhat common objects (as opposed to abstract figures known as tangrams). In response to both open-ended and specific questions during debriefing, none of the subjects reported suspecting that the Director's references were scripted.

Eyetracking apparatus

Matchers wore a headband-mounted visor fitted with an ISCAN RK-726PCI pupil/corneal reflection eyetracker and Polhemus head tracker. The visor supported a small video camera that outputted the Matcher's view of the referent array to the eyetracking computer. Another small video camera tracked the Matcher's eyegaze, which was overlaid as a cursor over the Matcher's view of the referent array (after being corrected by the head tracker for any head motion). The resulting video image was recorded on digital videotape. The apparatus weighed 6.5 oz. and enabled participants to move naturally while conversing in a seated position. The Directors' and Matchers' speech was input to the video recorder by an omnidirectional microphone.

Procedure

After giving informed consent, participants were introduced to the first confederate by the Experimenter. They were told that he was a research assistant helping with the experiment. Participants were then told that the purpose of the experiment was to study how they took direction from different people, and that a different research assistant would come in to give directions at several different times during the experiment. Participants were introduced to the task and were shown the cubbyhole display, along with two pictures on the first page in a loose-leave notebook that showed all the same objects but in different locations; one picture was labeled the starting arrangement and the other, the goal arrangement. The experimenter explained that the

Director would see the goal arrangement for each round in his notebook and then instruct the Matcher to move the objects around in the cubbyhole grid. Matchers were led to believe that this picture was all the Director would use to produce the instructions during the experiment. In actuality, each subsequent page in the Director's notebook contained not only the goal arrangement for a given round, but also prompts for all the scripted referring expressions (including hedges in Round 1) for both the objects to be moved and the landmarks in that round. During the experiment, the Director held the notebook at an angle in his or her lap so that the Matcher could not see it.

The experimenter told the Matcher and Director to communicate freely. He also explained that both the original and the new Director would have to wear sunglasses in order to prevent them from signaling the objects' locations through the direction of their gaze. Confederates were trained to gaze down at the notebook while delivering the first part of each direction (establishing the identity of the object to be moved) and then to keep the angle of their heads consistently aligned with the center of the grid. Once the Matcher touched the object, the Director reacted spontaneously. This partially scripted behavior appeared quite natural, as the Directors appeared to be looking down in order to determine the location of the object to be moved.

After a brief calibration procedure required by the eyetracker, the Director proceeded to direct the Matcher to move the objects. After each round was completed, the Director announced that the round was over, and then turned the page to the next goal arrangement. After turning the page following the third round, the Director read aloud a prompt from the notebook saying that it was time to go see if there was a partner change. He left the room and either he returned, or else the second Director entered and took over. After the fourth round, the new Director left and the original Director returned (if necessary). The Experimenter then removed the set of objects from the cubbyholes, put them away, and arranged the next set of objects in view of the partners.

Coding and data analysis

The digital video recordings of the critical rounds were used to code the time and position of the cursor representing the addressee's gaze superimposed over the scene camera, using a Sony DSR-30 digital video recorder with a shuttle knob that played the video frame-by-frame while shifting the audio pitch up to the normal range so that speech could be heard and easily understood. The temporal resolution of the image was 33.3 ms (corresponding to a video frame rate of 30 frames per second). The segments of interest began with the auditory onset of the first content word in a critical referring expression and continued until the addressee touched the target object. Coding began five frames before the segment of interest, in order to detect and exclude any cases where the addressee might happen to be already looking at the target by chance (rather than at the central fixation cross) before the critical expression was uttered.[2] A look was coded when the eye cursor remained within the same box for three or more successive frames (at least 100 ms, as in Barr & Keysar, 2002) and ended when the eye cursor left the box. Looks were categorized as being upon the central fixation cross, the box containing the target object, the box containing another object, or an empty box in the display grid. Errors were counted for the rounds during which the addressee touched an object other than the target object.

There were a total of 192 critical rounds (24 addressees $\times$ 8 sets of objects $\times$ 1 target object per set). Three critical rounds were missing from the data because their audio was not captured on the videotape, so a total of 189 rounds were coded.

We begin the results section by presenting the data on time to touch the target objects; in this interval the Matcher heard the critical referring expression, reached a hypothesis about what the Director meant by it, completed a decision process, and reached for the object. We were particularly interested in any partner-specific effects in the early moments of processing an old (original) or new referring expression, and so we present the time to the first look to the target object after the onset of the critical referring expression. Next, we examine patterns of looks to non-target objects in the display. For additional information on the decision process with which people mapped familiar versus new expressions onto objects, we measured the total number of looks to the target object before the touch. All measures were analyzed with repeated-measures ANOVAs crossing speaker (original, new) $\times$ expression (original, new). ANOVAs are reported in pairs, as F_1 (with subjects as the random factor) and F_2 (with items as random). By-subjects ANOVAs also included the between-subjects variable of Version (each subject experienced one of four lists in which critical rounds occurred in each of the four cells of the design). Planned contrasts (t_1 and t_2, by-subjects and by-items, respectively) were conducted for new expressions produced by the original versus the new speaker.

[2] On the whole, addressees followed the instructions to gaze at the fixation cross at the beginning of each round. Only twice was an addressee gazing at the target object at the onset of the critical expression. On one of these occasions, the addressee gazed away within 200 ms and so this was not counted as the first look to the target. The other time, the addressee continued to gaze at the target after hearing 200 ms of the critical expression; this was coded as a first look to the target at 200 ms past the onset of the critical expression.

Results

Upon hearing the original expressions, addressees touched the target objects equally quickly regardless of speaker (Table 2), replicating Barr and Keysar (2002). However, addressees processed the new expressions differently depending on who produced them. In the critical 4th rounds, it took addressees 540 ms longer to touch the target object when the new expression was produced by the original speaker than when the same new expression was produced by the new speaker, planned contrast, $t(23) = 2.30$, $p = .031$, $SE = 235$; $t_2(7) = 2.42$, $p = .058$, $SE = 239$. Speaker identity interacted with expression reliably by-subjects, $F_1(1, 20) = 9.47$, MSE = 2,510,845, $p = .006$ and marginally by-items, $F_2(1, 7) = 3.96$, MSE = 732,791, $p = .087$. The main effects of speaker and expression were reliable by-subjects and marginally-so or unreliable by-items [$F_1(1, 20) = 4.32$, MSE = 1,129,811, $p = .05$; $F_2(1, 7) = 4.87$, MSE = 452,252, $p = .063$, and $F_1(1, 20) = 21.74$, MSE = 3,836,201, $p < .001$; $F_2(1, 7) = 2.53$, n.s., respectively]. So addressees were slowest to map expressions onto referents when conceptual pacts they had reached previously with the same speaker were broken.

This pattern was even clearer early in processing. Addressees' very first looks to the target took 286 ms longer when new expressions were produced by the original speaker than by the new speaker, whereas (as Barr & Keysar, 2002, found) original expressions were processed just as quickly whether they were produced by the original or new speaker (Table 3). Speaker identity interacted with expression, $F_1(1, 16)$[3] = 8.86, MSE = 942,217, $p = .009$; $F_2(1, 7) = 7.011$, MSE = 259,359, $p = .03$. In the planned contrast, first looks were later upon hearing new expressions spoken by original speakers than by new speakers, $t(22) = 2.30$, $SE = 120$, $p = .03$; $t_2(7) = 2.42$, $SE = 124$, $p < .05$. There were no main effects for speaker, $F_1(1, 16) = .86$, n.s.; $F_2(1, 7) = 1.55$, n.s., nor for expression, $F_1(1, 16) = 1.35$, n.s.; $F_2(1, 7) = .68$, n.s.

One concern is whether an expression's given or new status and its resulting articulation could account for these differences. Previous research has found that both duration and intelligibility of repeated words are typically reduced in dialogue (e.g., Bard et al., 2000; Dahan, Tanenhaus, & Chambers, 2002; Fowler & Housum, 1987; Lieberman, 1963; Samuel & Troicki, 1998). Our original speaker had articulated each of the original expressions at least three times by the critical round, and so these expressions all specified given information, whereas from the perspective of the new speaker, all expressions specified new information (and she was not aware of which were new and which were given from the perspective of the addressee). If the new expressions were shorter or less intelligible when articulated by the original speaker than by the new speaker, this might account for the longer times to addressees' first looks and touches in that cell of the design. We measured the length of each critical referring expression and its adjective; these mean lengths are displayed in Fig. 2 along with the mean latency to the first look to the target. There was no difference in the duration of referring expression by new vs. original speakers, $F_1(1, 20) = .122$, n.s.; $F_2(1, 7) = .000$, n.s.,[4] nor was there any speaker-by-expression interaction, $F_1(1, 20) = .1.77$, n.s.; $F_2(1, 7) = .91$, n.s. Most important, for new expressions alone,

Table 2
Mean latency (and standard deviation) of target touch in ms

Expression	Partner	
	Original	New
Original	1897 (455)	1985 (401)
New	2595 (1343)	2055 (791)

Table 3
Mean latency (and standard deviation) of first look to the target in ms

Expression	Partner	
	Original	New
Original	1000 (557)	1038 (417)
New	1253 (465)	967 (777)

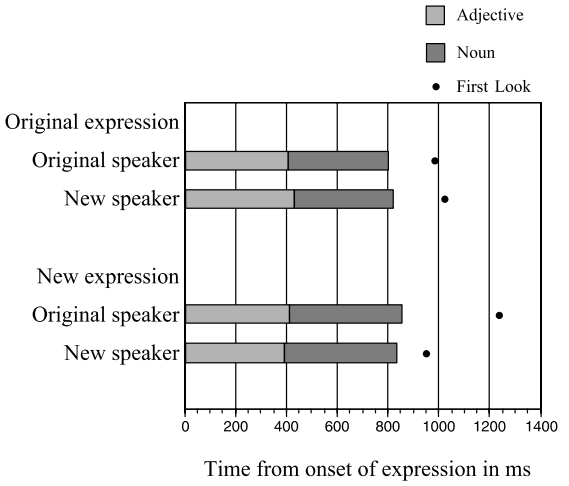

Fig. 2. Duration of critical referring expressions and onsets of the first looks to the targets.

[3] These degrees of freedom are 16 rather than 20 due to missing data from subjects whose eyegaze did not always reach the target object before they touched it.

[4] Comparing the lengths of original vs. new expressions is not informative here, as these constituted different lexical items.

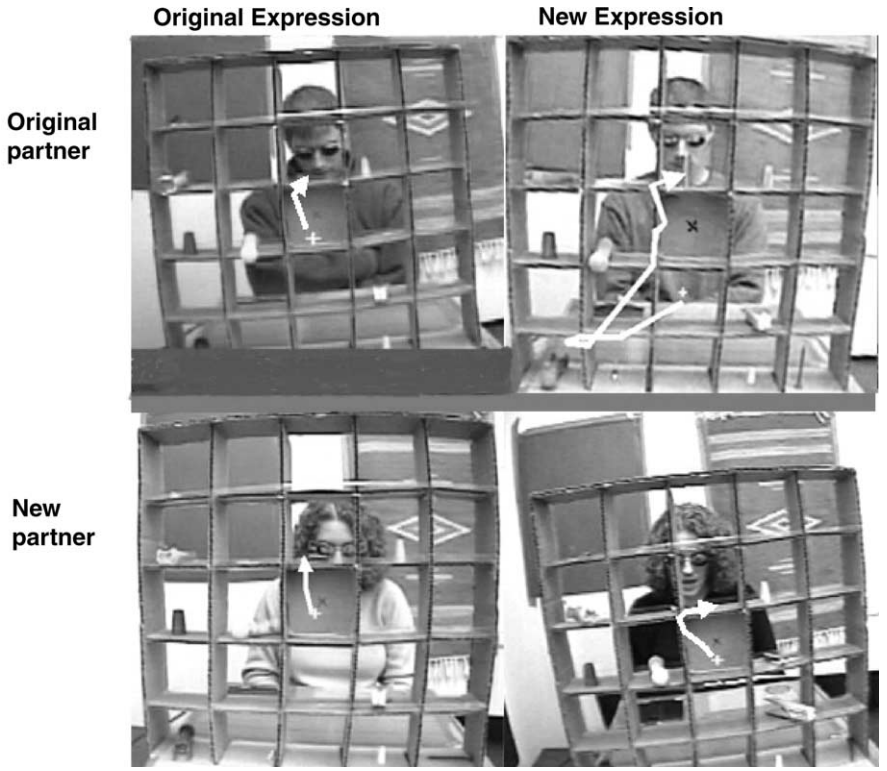

Fig. 3. Path of eyegaze of four different Matchers for references to the same object in all four conditions.

there was no durational difference by speaker, $t1(23) = .36$, n.s.; $t2(7) = .66$, n.s. Although we did not measure intelligibility for each token of the 8 original and 8 new critical expressions, the fact that original expressions were interpreted equally quickly whether produced by old or new speakers suggests that both speakers spoke intelligibly. The original expressions took 24 ms less to articulate when produced by the original speaker than by the new speaker, but this difference was not reliable, $t1(23) = .74$, n.s.; $t2(7) = .78$, n.s.[5] We will take up the issue of given- and new-marking later.

We now turn to the patterns of gaze toward non-target objects that preceded the touches and first looks to the target objects. When the original speaker abandoned the expression he had used previously, addressees appeared to search the display more thoroughly, as if to check for an entirely new object that had not yet been referred to. Fig. 3 shows representative paths of four different addressees' eyegaze superimposed over the display during critical rounds with the same target object from each of the four conditions, with the two confederates visible behind the display (the video clips from this figure can be downloaded from www.psychology.sunysb.edu/sbrennan/MBentrain). The upper right-hand quadrant shows a frame from a round in which the original speaker produced a new referring expression (*the silver pipe*). The addressee glances from the region of the fixation cross toward the lower left and upper left before fixating the target object, which is in the box just above the fixation cross. In the three other conditions, in which no conceptual pact is broken, gaze patterns show virtually identical, swift looks to the target. Table 4 shows that addressees looked at nearly twice as many other objects before looking at the target object when a conceptual pact was broken than in the other 3 conditions, speaker-by-expression interaction,

Table 4
Mean number of looks to non-target objects (and standard deviation) before the first look to the target object

Expression	Partner	
	Original	New
Original	1.111 (1.32)	1.125 (1.02)
New	2.167 (2.04)	1.125 (1.68)

[5] Note that this comparison was between identical tokens uttered by different speakers, whereas the usual comparison for given/new effects is between identical tokens by the same speaker.

$F_1(1,20) = 10.22$, MSE = 7.315, $p = .005$; $F_2(1,7) = 12.71$, MSE = 2.214, $p = .009$.[6]

We next examined looks to some of the non-target objects in more detail, post-hoc.[7] Although we did not systematically pre-designate specific competitors in these sets, recall that 6 of the 8 sets of objects contained a non-target object that shared an attribute with the target object such that it *could* also have been mapped onto the adjective that was used to begin the new critical referring expression for the target object (e.g., in the set with *the shiny cylinder/silver pipe*, there was also an object consistently called a *big hook* that also happened to be silver). To the extent that these similar objects may have attracted more attention than less similar objects in the array as candidates for mapping to the new critical referring expressions,[8] we expected that addressees might gaze at them somewhat differently depending on who the speaker was. That is, when the new speaker produced a new expression, the addressee should spend more time before the touch considering those objects in the display that were especially good mappings to the expression (e.g., upon hearing the adjective *silver*..., deciding between *the shiny cylinder* and *the big hook*). However, when the original speaker used a new expression, inexplicably breaking a conceptual pact, we expected the addressee to launch a more thorough search of the entire display, perhaps looking for an object hitherto overlooked. With the original speaker, the addressee might spend less time considering the object known as *the big hook* because mapping *the shiny cylinder* onto that object would require breaking yet another conceptual pact with that speaker. The pattern of means supported this idea: On average, addressees gazed at the similar objects for 128 out of 2055 ms before the touch to the target object when the new speaker uttered the new expression and 79 ms out of 2595 ms before the touch when the original speaker uttered it. The proportions of time before the touch spent looking at the similar objects were reliably different by items only, $t_1(11) = 1.62$, SE = 94.2325, $p = .13$; $t_2(5) = 2.66$, SE = .005981, $p < .05$ (with only 6 items, the power of these exploratory analyses was .32 and .57, respectively).

What is particularly striking is that addressees were so fast to initially look at and eventually touch target objects upon hearing *new* speakers produce entirely new expressions. In fact, *both* original and new expressions produced by new speakers resulted in equally fast first looks and touches to the target as did the original expressions produced by the original speakers (Tables 2 and 3). This may have occurred for two reasons: (1) The new and original expressions mapped to equally "good" perspectives on the target objects, such that the new expressions also provided strong cues that were easily mapped to the target objects, and (2) addressees had encoded the referent arrays extremely well by the time they heard these expressions (by Round 4, each object had been referred to at least three times previously, the addressees had scrutinized and handled each object enough to become quite familiar with its characteristics, and they had personally placed each one in its current location).

To judge whether this explanation was plausible, we returned to the videotaped data and coded the time to first look for Rounds 2 and 3.[9] Note that it is a robust finding in referential communication studies that the time to complete a trial, number of words, and errors all decrease during repeated referring for the first several repeated trials (see, e.g., H. Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989; Schober & Brennan, in press) and that these measures asymptote as partners approach maximal efficiency in referring. For original expressions uttered by original speakers, we contrasted the times to first look to the target for Round 2 vs. 3 and 3 vs. 4. Round 3 was 464 ms faster than Round 2, reliable by-subjects, $F_1(1,17) = 23.61$, MSE = 4,914,934, $p < .001$ and marginally so by-objects, $F_2(1,7) = 5.19$, MSE = 1,845,940, $p = .06$. Round 3 was slightly but not reliably slower than the comparable Round 4s (in which the original speaker used the original expression), $F_1(1,17) = 3.87$, MSE = 519,287, $p = .07$; $F_2(1,7) = 1.14$, n.s. This is consistent with the idea that by Round 3 addresses were extremely familiar with the objects in the display and were reaching a ceiling on the efficiency with which they mapped expressions to referents.

An overall difference in Round 4 between processing brand-new versus familiar referring expressions did, however, emerge as a main effect in the *number of looks* addressees made to the target object before touching it. Addressees made, on average, just under one look to the target upon hearing familiar expressions, but more than one look upon hearing new expressions, regardless of which speaker produced the expressions (Table 5), different at $F_1(1,20) = 11.36$, MSE = 2.50, $p = .003$; $F_2(1,7) = 18.39$, MSE = .845, $p = .004$. In our task, addressees had already searched the display, so reexamining an object after hearing the critical expression suggests a decision process during which the object was considered more than once; perhaps after initially looking at a target object, addressees had more difficulty deciding to map the referring expression onto it when

[6] This interaction led to marginal effects of speaker ($F_1(1,20) = 5.55$, MSE = 5.753, $p < .03$; $F_2(1,7) = 3.81$, MSE = 2.217, $p < .10$) and expression ($F_1(1,20) = 6.34$, MSE = 6.253, $p = .02$; $F_2(1,7) = 2.71$, MSE = 1.975, $p < .15$).

[7] We thank an anonymous reviewer for this suggestion.

[8] We do not call these similar objects competitors, as we did not norm the target expressions to see if they were equally good for these objects.

[9] Round 1 was not included in this analysis because those references often included hedges, and so the referring expressions were not exactly the same as those in subsequent rounds.

Table 5
Mean number of looks to the target (and standard deviation) prior to touching it

Expression	Partner	
	Original	New
Original	.93 (.58)	.96 (.36)
New	1.27 (.76)	1.27 (.74)

the expression was new than when it was familiar. There was neither a main effect of speaker ($F_1(1,20) = .02$, n.s.; $F_2(1,7) = .01$, n.s.) nor an interaction of speaker with expression ($F_1(1,20) = .01$, n.s.; $F_2(1,7) = .02$, n.s.) for the number of looks to the target before the touch.

Errors were recorded whenever an addressee touched a non-target item on a critical round. Only 13 errors were committed during the 189 recorded rounds. Although there were not enough errors to conduct an analyses of variance, the pattern is worth noting. Nine errors occurred with new expressions and four with original expressions. Six of the 13 errors occurred when the original partner used a new expression; this was twice as many errors as in any other condition.

Discussion

This study focused on what happens when a conceptual pact is broken—in particular, how an addressee processes a referring expression when it departs from a precedent set with a particular speaker. On the pragmatic level, the Maxim of Manner suggests that expressive choices by speakers license implicatures by addressees (Grice, 1975), and in particular, the Principle of Contrast predicts that "speakers take every difference in form to mark a difference in meaning" (E. Clark, 1993, p. 69; see also E. Clark, 1997). That is, when a speaker abandons a precedent, there should be a discernable reason. Speakers often have, of course, good reasons to change terms; for instance, if the figure-ground relationship for a referent changes, they may need to choose more specific referring expressions in order to be sufficiently informative (Brennan & Clark, 1996; Brown, 1958a; Olson, 1970). In the current experiment, there was no such figure-ground change to cue or motivate an abandoned precedent, but only a change in speaker (for half of the abandoned precedents). We found that addressees were slower to process new expressions (that departed from the expressions that they had already entrained upon with a particular speaker) when the new expressions were uttered by the original speaker than by a new speaker. As for the original (already-entrained-upon) expressions, addressees processed these just as quickly when they were uttered by the original speaker as when they were uttered by a new speaker who was not co-present to the previous use of the expressions.

These partner-specific differences emerged early in processing—not only by the touch (with a difference of 540 ms), but by the addressee's first look to the target object (with a difference of 286 ms). Therefore it makes sense to consider the underlying representation in memory that could support this effect. We propose that the representations from which entrainment in conversation emerges include not only an episodic association between a referent and a referring expression (with an associated perspective), but also relevant information about the context of use, which in this case includes speaker-specific information. When the addressee heard an entrained-upon expression, this acted as a strong cue for the mapping previously associated with that expression, regardless the speaker. An entirely new expression, however, cued no such mapping, and so a new one was forged. A new expression uttered by a new speaker provided a compound cue for which both parts of the cue were new (see Gerrig & McKoon, 1998; Myers & O'Brien, 1998); because the new expression cued no existing mapping, the addressee had to determine which object best qualified for the new mapping. In our experiment, this was done with relative ease; the new expression provided a good cue to the features of the target object, all the objects were extremely well-learned (having been discussed and positioned by the addressee just beforehand), and some background objects had been eliminated already by virtue of having been referred to using other expressions. And because the original speaker was absent when the new speaker was present, it is possible that the original referent-expression mapping may have decayed somewhat, making it less likely to interfere with a new mapping involving the same referent (see Altmann & Gray, 2002).

In contrast, when new expressions were uttered by the original speaker, the compound cue was less good for the addressee. The process of making a new mapping from the target to the new expression was slowed, we propose, not only by the fact that the original speaker cued a different referring expression for that target, but also by the fact that any alternative objects under consideration for a new mapping were already strongly associated with different referring expressions. In the *original-partner, new expression* cell of our design, addressees sometimes appeared to launch a wide search over the display, as if to check for an unfamiliar object that they might have thus far overlooked. They produced nearly twice as many looks to non-target objects before their very first look to the target object as in any of the other cells (Table 4).

Despite the cost of abandoning a precedent set with the original speaker, there was surprisingly little cost to forging a new mapping when necessary with a new speaker (at least in this simple world, with the ample perceptual support provided by our task). By the critical rounds, the objects were already well known, and the addressee had placed every one of them in their current

locations. With the new speaker, the time to the first look and the time to the touch were about the same for new expressions as they were for old expressions. Nevertheless, a difference emerged in the processing of new vs. old expressions with respect to the number of looks to (and subsequently, away from) the targets. Upon hearing new expressions, addressees produced slightly but reliably more looks to targets before touching them, *regardless of speaker*, suggesting more uncertainty in making the decision to map new expressions to these referents (Table 5). Interestingly, this apparent uncertainty resulted in little or no time cost when the new expression came from a new speaker.

Conclusions

This study presents the first on-line evidence for a speaker-specific effect early in the interpretation of referring expressions by addressees. It also demonstrates that referring is flexible; even after addressees have entrained on expressions and formed a conceptual pact with one speaker, they can easily take up new expressions and reach new conceptualizations of the same objects with another speaker. They can flexibly map new expressions onto familiar referents without interference, unless it is the *original* speaker who has introduced a new expression, inexplicably breaking a conceptual pact; this results in a small but measurable delay, both in the early moments of the resolution process and in the subsequent decision process before touching the object.

As we noted earlier, our findings are consistent with Barr and Keysar's (2002), even though our conclusion is not. Barr and Keysar concluded from their experiments that partner-specific knowledge of conceptual precedents does *not* figure in the interpretation of referring expressions. However, their Experiment 2 tested only half of our design; they found no effect of original vs. new speaker on the interpretation of expressions that had been previously entrained-upon with the original speaker. Our results for that half of the design concur with theirs. Our experiment added the cells in which the original speaker breaks a conceptual pact, inexplicably abandoning an entrained-on expression in order to use a new, equally-specific expression, compared to when a new speaker uses that same new expression. We found evidence for a partner-specific effect in the form of interference, rather than the facilitation sought in Barr and Keysar (2002)'s Experiment 2.

In Barr and Keysar's Experiment 3, either an original or a new speaker (both prerecorded) switched terms after having spoken the same subordinate expressions five times (e.g., switching to *car* after having repeatedly used *sportscar)* while listeners selected from only two referents that happened to be members of the same acoustic cohort (e.g., *car* and *carnation*). So the target expression (*car*) was new for the target picture, whereas the foil expression (*carnation*) had been used five times to refer to the foil picture, creating a strong bias toward the latter mapping for the cue *car*... That the experiment found no speaker-specific difference in listeners' tendencies to look at the carnation is not particularly surprising; the bias for *car* to cue the well-learned association with carnation may well have been strong enough to obscure any effect of a partner-related cue.

The finding that addressees interpret utterances differently depending on who says them is consistent with well-known pragmatic principles such as the Maxim of Manner (Grice, 1975) and the Principle of Contrast (E. Clark, 1987, 1997). It is generally assumed that pragmatic effects in interpretation result from strategic or inferential processes and are relatively slow to emerge (e.g., Keysar, Barr, Balin, & Paek, 1998). But the present demonstration of a partner-specific effect with the addressee's very first look to the target object suggests that certain pragmatic processes may also have an automatic component—supported by episodic representations that can have an impact relatively early in processing.

Precisely what serves as the partner-specific or context-specific cue for these representations is not yet clear, and suggests directions for further study. A representation could be tagged, perhaps abstractly, with the speaker's identity, or there could be a perceptual basis for the cue, such as the speaker's voice or physical presence. Alternatively (or in addition), partner-specific cueing could emerge if a speaker marks an expression prosodically, as given or new. In an experiment by Dahan et al. (2002), pre-recorded accented and de-accented referring expressions from a single speaker were played to listeners, and so could be precisely controlled such that each listener heard the same tokens. This experiment found a bias on the part of listeners to interpret de-accented expressions as anaphoric (indexing a previously mentioned or given referent) and accented expressions as new. In our experiment, all expressions were new for the new speaker. As for the original speaker, his new expressions were somewhat longer than his original (given) expressions, but not reliably so. So our findings do not address whether the nature of the partner-specific cueing was due to the speaker's identity, physical presence, voice, or history with the expression (as signaled by given/new marking). We did not control these things because we wished to examine interpretation within an interactive conversational setting; each speaker attempted to produce the scripted expressions similarly, in as natural and felicitous a means as possible. To tease these possibilities apart would likely require playing pre-recorded instructions to listeners, as in Dahan et al.'s (2002) methodology.

Finally, we wish to emphasize that a delay that may ensue from speaker-specific interference need not be problematic for comprehension; first of all, the delay is extremely short, and second, it may in fact provide an

advantage. The fact that pragmatic or speaker-specific expectations can constrain the resolution process in this way points to a system that is ready to notice contextual shifts (such as a change in the speaker's goal or in the set of objects a referent needs to be distinguished from) when a speaker breaks a conceptual pact. At the same time, since there are so many possible ways to label or describe most objects in the world, people in conversations with new partners need to be able to ignore precedents established with previous partners. This sort of flexibility in the mapping of expressions to referents distinguishes natural languages from formal languages; it is crucial that referring be flexible in order for human communication to work as well as it does.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants No. 0082602 and 9980013. We thank Darron Vanaria for his assistance and Gregory Zelinsky, Richard Gerrig, Joy Hanna, Sid Horton, Nancy Franklin, Michael Schober, Martin Pickering, Ellen Bard, and Herb Clark for helpful discussion and comments.

Appendix

Original and new expressions for target items

Original expression	New expression
Blueprint car	Yellow car
Plastic hook	White hook
Table lego	Multicolored bridge
Steel wool	Wire ball
Roof top	Paper tent
Slinky	Loose wire
Space man	Blue guy
Shiny cylinder	Silver pipe

References

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science, 13*, 27–33.

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., & Keysar, B. (2002). Anchoring and comprehension in linguistic precedents. *Journal of Memory and Language, 46*, 391–418.

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. (1958a). *Words and things*. Glencoe, IL: The Free Press.

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

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, JN: Erlbaum.

Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Clark, E. V. (1997). Conceptual perspective and lexical choice in acquisition. *Cognition, 64*, 1–37.

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

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47*.

Fowler, C., & Housum, J. (1987). Talkers' signalling of 'new' and 'old' words produced in various communicative contexts. *Language and Speech, 28*, 47–56.

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

Gerrig, R. J., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes, 26*, 67–86.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3: Speech acts* (pp. 225–242). New York: Seminar Press.

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.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*, 172–187.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26*, 131–158.

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

Samuel, S. G., & Troicki, M. (1998). Articulation quality is inversely related to redundancy when children or adults have verbal control. *Journal of Memory and Language, 39*, 175–194.

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