How Toddlers and Preschoolers Learn to Uniquely Identify Referents for Others: A Training Study

Danielle Matthews
University of Manchester, United Kingdom

Elena Lieven
University of Manchester, United Kingdom, and Max Planck Institute for Evolutionary Anthropology, Germany

Michael Tomasello
Max Planck Institute for Evolutionary Anthropology, Germany

This training study investigates how children learn to refer to things unambiguously. Two hundred twenty-four children aged 2.6, 3.6, and 4.6 years were pre- and posttested for their ability to request stickers from a dense array. Between test sessions, children were assigned to a training condition in which they (a) asked for stickers from an adult, (b) responded to an adult's requests for stickers, (c) observed 1 adult ask another for stickers, or (d) heard model descriptions of stickers. All conditions yielded improvements in referring strategies, with condition (a) being most effective. Four-year-olds additionally demonstrated learning effects in a transfer task. These results suggest that young children's communication skills develop best in response to feedback about their own attempts at reference.

Human communication depends fundamentally on our ability to draw other people's attention to things by referring to them. Most infants begin to do this around their first birthdays when they point to things in order to request them or simply to inform other people of them (Liszkowski, 2006). With the advent of language, a whole new array of referring strategies is opened up and children become increasingly skilled in choosing the right strategy for the situation at hand. Thus, on one occasion an object may be identified simply by pointing at it and using a relatively uninformative referring expression like ''That one!''. On other occasions, the object may be less accessible for the addressee, in which case complex referring expressions may be called for, like "the funny looking one with the green handle up on the top shelf" (Ariel, 1988; Gundel, Hedberg, & Zacharski, 1993). The question that we address in this paper is precisely how children learn to use such sophisticated referring strategies.

Ever since the 60s, children's ability to effectively refer to things has been assessed with various forms of the referential communication task (Glucksberg, Krauss, & Weisberg, 1966; Krauss & Glucksberg, 1969). Typically, this task requires the child to uniquely identify one object in a complex array of similar objects for another person. Original studies found that 5-year-olds were relatively poor at this task and indeed subsequent research has confirmed that young children's spontaneous object descriptions are often not informative enough to uniquely identify any one referent for the addressee (see Dickson, 1982, for a review). Importantly though, Deutsch and Pechmann (1982) established that, although first communicative attempts are often insufficiently informative, children from the age of 3 upward are generally capable of arriving at an adequate description of an object over several conversational turns. Thus, although 87% of 3-year-olds in their study requested toys in an ambiguous manner at first, when asked for clarification by the experimenter, 89% arrived at a satisfactory description. On the basis of these results, Deutsch and Pechmann questioned whether this process of breakdown and repair might in fact be a precondition for communicative development. That is, does attempting reference, failing, and receiving feedback ultimately lead children to become more efficient communicators?

A number of training studies with school-age children suggest that communicative breakdown and repair is sufficient to improve performance on referential communication tasks. For example, Robinson

and Robinson (1985) found that 5-year-olds performed better on referential communication tasks if they had been trained by an experimenter who did not act on the basis of the child's ambiguous object descriptions and simply told them ''I can't really choose [which object the child was talking about] yet.'' Providing an additional explicit explanation as to why the child's message was uninformative did not further aid learning in these conditions where communicative breakdown was marked by the behavior of the experimenter. However, in conditions without such noncompliant behavior to mark breakdown (where the experimenter successfully guessed at an interpretation of the ambiguous message but then explained to the child that his/her message had not been informative enough and why), children did not make comparable gains at posttest. It would thus appear that experiencing breakdown and repair is effective even in the absence of an explicit explanation of what has happened and furthermore simply explaining that a message is uninformative is not on its own as effective in improving communication.

Other training studies have shown that school-aged children improve on referential communication tests if they are (a) encouraged to focus on what is different about the referent in comparison to the other objects in view (Asher & Wigfield, 1981; Lefebvre-Pinard & Reid, 1980), (b) given adult models of how to describe referents (Whitehurst, 1976; Whitehurst, Sonnenschein, & Ianfolla, 1981), and (c) trained to assign blame for breakdowns in communication and given feedback on blame assignment (Sonnenschein & Whitehurst, 1984). However, despite the considerable improvements made by children in these training studies (Sonnenschein & Whitehurst [1984] trained 5-year-olds to the level of performance of typical 9-year-olds), few would claim that the training provided in these studies is representative of how children normally learn. Rather, these studies were designed to establish which component skills are required for successful communication—a question that had initially been overlooked in the rush to attribute communicative failure entirely to the monolithic construct of childhood egocentrism (Dickson, 1982). Consequently, the question still remains as to how far experiencing more naturalistic communicative breakdown and repair can help children's communicative skills develop.

Another important question concerns the age at which children might first benefit from the experience of repairing unsuccessful attempts at reference. The above-reviewed studies have focused on children aged 4 and older, so we do not know whether 2-year-olds, for example, would also benefit from

feedback. There is now ample evidence that children repair their failed communicative attempts even at the preverbal stage (Golinkoff, 1986, 1993; Liszkowski, Carpenter, & Tomasello, in press; Tomasello, Carpenter, & Liszkowski, in press) and often respond to clarification requests in an appropriate manner by the age of 2 (Anselmi, Tomasello, & Acunzo, 1986; Gallagher, 1977; Tomasello, Anselmi, & Farrar, 1984/ 1985). However, it is not clear that children this young are able to infer from the fact that someone has asked for clarification that their original attempt at reference was not sufficiently informative (as opposed, e.g., to not being heard). Thus, it is possible that children are able to respond to clarification requests without fully understanding why those requests were made—that is, without questioning why their original attempt had not succeeded. In fact, many studies show that attributing the cause of referential failure to the lack of information in a prior message is a skill that is particularly slow to develop (Lloyd, Mann, & Peers, 1998; Robinson & Robinson, 1985). It is therefore unclear whether repairing utterances in response to clarification requests would have an effect on young children's subsequent attempts at communicating.

There are very few studies that have investigated the effect of feedback on communication. Robinson and Robinson (1981) found that 6-year-olds were better able to detect message ambiguity if, as preschoolers, their mothers had explicitly told them when they could not understand them (as assessed by coding naturalistic recording made of these same children between the ages of 2 and 3). The only studies to our knowledge that have tested the effect of feedback with younger children were conducted by O'Neill and Topolovec (2001). In these studies, 2-year-olds saw a sticker dropped into one of two containers while their parent had their eyes closed. In one condition, the two boxes were far apart, whereas in the other the boxes were adjacent, making a pointing gesture toward them ambiguous. When children aged 2 years 8 months attempted to inform their parent of the location of the sticker, they were significantly more likely to name the box if the two boxes were adjacent than if they were far apart, suggesting some ability to adapt referential strategies to the communicative context (not observed at 2 years 4 months). Of interest for our current purposes were two feedback trials conducted at the end of the main experimental session. In these trials, if the child failed to verbally identify one of the adjacent boxes, the experimenter asked which box the child meant. The inability of some 2-year-olds to respond appropriately to this feedback suggested they were unaware of the referential ambiguity of their pointing gestures. Furthermore, the lack

of improvement on the second of these trials compared to the first led the authors to conclude that the children had not learnt from this feedback. This of course casts doubt on the idea that 2-year-olds appreciate why clarification requests are made and thereby learn from them. Nonetheless, this study had only one trial in which to assess learning. With more feedback and test trials, it is possible that these children would have become more informative. The purpose of the current study was to test whether children aged between 2 and 4 years of age would draw from the experience of needing to repair failed communicative attempts and become better communicators as a consequence.

In addition to testing whether children would benefit from receiving feedback about their own communicative attempts, we also tested whether experiencing communicative breakdown from the addressee's perspective and from an onlooker's perspective would drive learning, as previous studies have shown to be the case for older children (e.g., Lefebvre-Pinard & Reid, 1980; Whitehurst, 1976). Thus, in the current study, we test whether 2-, 3-, and 4-year-olds will benefit from being trained (a) in the *speaker* role with an experimenter giving feedback to the child as necessary with clarification requests, (b) in an *addressee* role where an experimenter-speaker sometimes refers to objects ambiguously and the child must attempt to understand, (c) in an *onlooker* role where the child observes an experimenter-speaker refer to objects (sometimes ambiguously) for an experimenter-addressee who attempts to understand and asks for clarification where necessary, and (d) in a *model description* condition where children are simply given models of appropriate object descriptions by an experimenter. This fourth condition allows us to establish whether it is the experience of breakdown and repair that drives communicative development or whether hearing relevant descriptive phrases alone would be sufficient. It is therefore intended to be a control condition that isolates the effect of training through feedback and repair (present in conditions a to c) while ensuring that participants in this fourth condition become equally familiar with the experimenters and task as in the other conditions (c.f. Asher & Wigfield, 1981). By providing model adult descriptions of the referents, this condition essentially demonstrates to the child ''how we talk about these stickers,'' thereby strengthening vocabulary and grammar relevant for the task.

In order to make the training study accessible and enjoyable for 2-year-old children, we created a sticker book task. At pre- and posttest, children needed to observe with one experimenter which sticker was missing in their book (compared to a completed book) and ask a second experimenter for the sticker they needed from an array of similar stickers. Training also involved making sticker books although the role of the child in this process varied according to experimental condition. Over the course of 3 days, children participated in a pretest, four training sessions, a posttest, and a transfer task. Our aim was to assess how children's referring strategies changed from pretest to posttest as a function of the training condition they were assigned to. In addition we wanted to assess children's performance on a transfer task both as a function of the current training conditions and as compared to data obtained in a previous study (Matthews, Theakston, Lieven, & Tomasello, 2006) that tested children of the same age on the same task but with no prior training of any sort.

Method

Participants

About 224 normally developing, monolingual, English-speaking children were included in the study (102 boys, 122 girls). There were sixty-nine 2-year-olds (range 2.1 – 2.11, mean age 2 years 7 months), eighty 3-year-olds (range 3.1 – 3.11, mean age 3 years 6 months), and seventy-five 4-year-olds (range 4 – 4.9 years, mean age 4 years 5 months). A further 41 children (seventeen 2-year-olds, twenty-three 3-year-olds, one 4-year-old) were not included either because (a) they did not want to complete or did not respond at all in the pretest (12 children), (b) they were absent for either the second or the third day of the study (12 children), (c) of parental intervention (one child), or (d) of experimenter error (16 children). The children were tested in the Max Planck Child Study Centre, Manchester, or in a quiet area in their nursery or primary school in the Manchester area, UK. Approximately 90% of the children were White European, with the largest minority ethnic group being Asian. All children were middle class, most attended part-time day care, and all 4-year-olds attended school full time.

Materials and Design

Six picture books were made about the adventures of a family (e.g., *The Bumbles Go to the Farm*, *The Bumbles Have a Party*). Each book was made up of six unrelated pictures in which different members of the family (mother, father, girl, boy) performed simple actions. Three pictures depicted intransitive scenes (e.g., the little girl crying) and three transitive scenes (e.g., the mum eating an ice-cream). Agents in all scenes were

animate. Of the three characters being acted upon in the transitive pictures, one was human, one was an animal, and one was inanimate. The actions depicted could all be described by verbs that are attested as frequent in the English child-directed speech made available on the CHILDES database (MacWhinney, 2000). A description of the pictures in all six books can be found in Appendix A.

For testing, each book was printed in two versions: the experimenter's copy, which was the original copy of the book, and the child's copy, which was identical to the experimenter's book except that all the characters were missing (Figure 1). The child's task was to make their book the same as the experimenter's by placing the relevant sticker onto the background of each page. Boards displaying 14 stickers were produced ensuring that, for any given test picture, there would be one matching sticker, one sticker that matched the character(s) but not the action they were performing, and one sticker that matched the action but not the character(s) performing them.

Procedure

Snap game. Before the main testing session began, the child and Experimenter 1 (E1) sat together at a table and played a simple "snap" game. In this game, six pairs of identical cards were shuffled and spread out on the table. E1 took one card and asked the child if s/he could find the other card that was "the same as this one." The game continued in this fashion until all six pairs of cards had been matched together. This task served to ensure that the children understood the term "the same" and would be able to make their incomplete book the same as E2's book.

Introduction and pretest. After the snap game, E1 went to sit on the other side of the room and E2 began the pretest with the child. To introduce the task, E2 came over to the table and presented the two versions of the picture book to the child, explaining one was for her and the other was for the child. E2 explained that she had finished making her book that morning but that child's book was not quite finished and it had some pictures missing. E2 proposed that the child could finish making his/her book and take home if s/he wanted to. To ensure that all the referential terms required were familiar to the child, E2 asked him/her to point to each of the characters on the cover of the picture book in turn, which all the children could do (Figure 1). E2 encouraged the child to name each character but did not insist if s/he did not want to. E2 explained that some pictures would be missing from the child's book and that E1 had lots of stickers that the child could use to finish the book and make it exactly the same as E2's book. E1 had meanwhile moved to the opposite side of the room where she had a collection of stickers fixed to a board on the wall. E2 made it explicit to the child that from where E1 was sitting she could not see their books because there was a large box on the table that blocked her view. E2 turned the cover pages of her and the child's book over to reveal the first picture in its complete and incomplete versions (Figure 1). She asked the child if s/he could go and get the sticker from E1 that would make their books the same.

E1 encouraged the child to come and retrieve the sticker s/he needed. Once the child was in front of the sticker board, both the child and E1 could easily see all the stickers. However, the stickers were fixed sufficiently high on the wall so that any pointing gesture on the part of the child would not uniquely identify any given referent. An overview of the experimental set up is given in Figure 2.

For the purposes of the pretest, all the children were required to collect six stickers in turn to complete their picture book. The children varied as to how
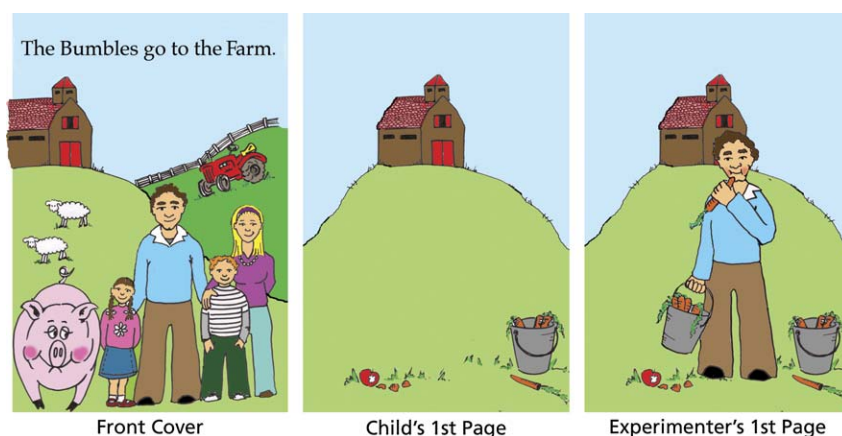

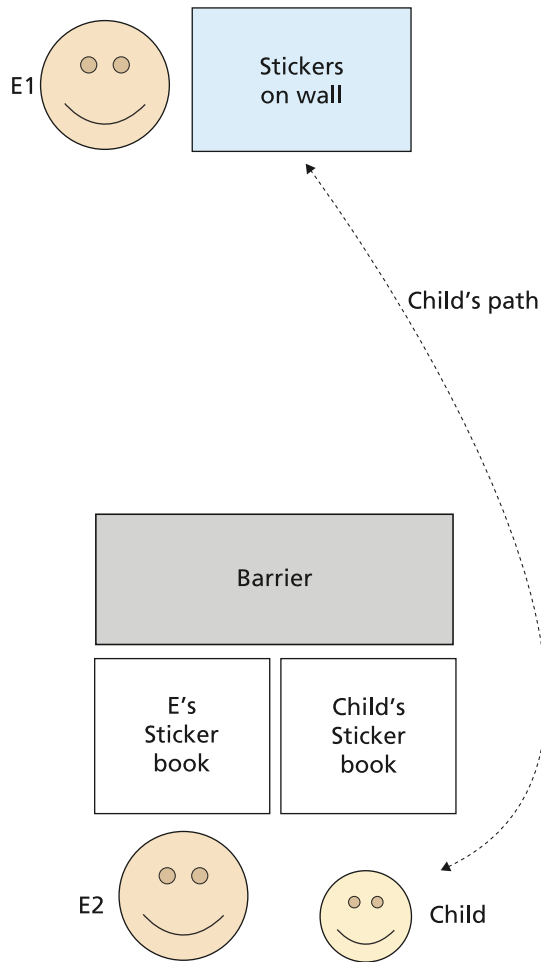*Figure 1.* Example pages from books used in the sticker task.

Figure 2. Layout of room for pre- and posttest.

they attempted to do this. If the child fully described or simply named the character in the sticker, then E1 handed over the correct sticker. If the child only pointed and/or said "that one," then E1 first selected an incorrect sticker and asked "this one?." If the child rejected the incorrect sticker, E1 next selected the correct sticker and the child returned to the picture book with it. Occasionally, however, the younger children would accept the first, incorrect, sticker. In this case, E1 said, "ok here it is." When the child returned to the books, E2 then pointed out to the child that the incorrect sticker did not make their books the same and so the child should return to E1 and get the right sticker. (A total of twenty-seven 2-year-olds and eighteen 3-year-olds accepted a wrong sticker at least once in this fashion. Only two did so as many as three times.) Similarly, if children explicitly asked for an incorrect sticker (e.g., asked for Father Christmas when they needed the daddy), they were handed an

appropriate sticker and E2 pointed out that it was incorrect in the same way. (A total of twenty-nine 2-year-olds, eleven 3-year-olds, and three 4-year-olds asked for an incorrect sticker on at least one occasion. Only two did so more than twice, none more than four times.)

None of the children were given any other feedback in the pretest.

Training sessions. Each child was randomly assigned to one of the four between-subjects training conditions described below. Training took place over four sessions, spread over 3 separate days. There was a maximum of 8 days between the first and the final session. The first training session took place immediately after the pretest. A further two training sessions were carried out on the next training day. The final training session and posttests were carried out on the third training day. Each training session consisted of the completion of a new book. The child's task during training differed according to condition as follows.

In the *child-is-speaker* condition, the task for the child was to request stickers from E1. This task was similar to the pretest except that E1 provided feedback for the child if s/he did not uniquely identify the sticker needed. If the child only pointed at the sticker, then feedback consisted of informing the child that E1 could not see which sticker s/he needed and asking the child to *tell* her which one she needed. If the child referred to the character ambiguously (e.g., "the girl one"), then E1 would ask for clarification so as to uniquely identify the sticker (e.g., "the girl eating an ice-cream or the girl swimming?"). If the child uniquely identified the sticker, E1 located it on the board and handed it over immediately.

The *child-is-addressee* condition differed to the speaker condition above in that E1 and the child reversed roles. Thus, in the transition from pretest to training, E1 explained that it was her turn to make a book and asked the child if she would give her the stickers she needed. The layout of the room was the same as in Figure 2 except that the child and E2 had a set of stickers laid out on the table in front of them and E1 had two books on the other side of the room (not visible to E2 and the child). E1 looked at the books on her side of the room and then came over to the child and asked him/her for the necessary sticker. The child's task was to hand it over.

E1 was only occasionally informative in her requests. For two of the six stickers, E1 simply requested "that one" and pointed to the sticker board. If the child requested clarification or looked back blankly, then E1 proceeded to describe the sticker more adequately (e.g., "oh sorry, I need the girl swimming"). If the child randomly chose a sticker, then

E1 pointed out that it was the incorrect sticker and then informatively requested the correct sticker (e.g., "Sorry not that one. The girl singing"). For another two stickers, E1 only named the character (e.g., "I need the girl"). If the child requested clarification, then E1 subsequently resolved the ambiguity (e.g., "the girl singing"). If the child directly handed over a sticker, E1 said this was the wrong sticker and clarified which alternative she needed (e.g., "Sorry not that one. The girl singing"). So, if the child ever guessed which sticker was needed on the basis of too little information, s/he was always told that the guessed sticker was not the right one and was given a full description of an alternative sticker. For another two stickers, E1 uniquely identified the referent and the children always identified the correct sticker.

The *child-is-onlooker* condition was like the addressee condition except that E1 asked E2 for stickers instead of asking the child. As in the child-is-addressee condition, E1 was only occasionally informative. For two of the six stickers, E1 simply requested "that one" and pointed to the sticker board. E2 consequently requested clarification by asking "Which one?" and E1 proceeded to describe the sticker in full. For another two stickers, E1 only named the character. In this case, E2 requested clarification by asking, for example, "Which one, the girl crying or the girl singing?" and E1 subsequently resolved the ambiguity by describing the correct sticker in full (e.g., "the girl singing"). For another two stickers, E1 uniquely identified the referent and E2 understood perfectly. To maintain interest, the child acted as an intermediary picking up the stickers from E2 and handing them over to E1. Thus, in this condition the child was required neither to make nor to interpret a request but simply observed others negotiating reference.

In the *model description condition*, E1 came to sit at the table with E2 and the child. E1 sat with a set of stickers across a table from E2 and the child, who had their picture books. Everyone could see the picture books and the stickers (the barrier was removed from the table). As the child and E2 turned the pages, E1 would describe which sticker the child needed and immediately hand it over (e.g., "Ah, you need the girl singing. Here you are."). The task for the child was to complete the book, just as was the case in the child-is-speaker condition except that no verbal requests were necessary. Thus, children in this condition were exposed to informative adult descriptions of each sticker but observed no negotiation of reference be this from a first-, second-, or third-person perspective.

Posttest. After the final training session on the third training day, the child completed the posttest and then a transfer task. The posttest took exactly the same form as the pretest described above. To minimize item effects, half the children had *The Bumbles at Christmas* as their pretest and *The Bumbles Have a Party* as their posttest; the other half had the opposite arrangement. The remaining four storybooks were used for the training sessions.

Transfer task. The transfer task was based on a study (Matthews et al., 2006) that tested children's pragmatic understanding of reference but differed considerably from the sticker book task in that (a) it did not involve requests and (b) the referents in question were not visually available to the child's addressee. For this task, the child was shown a video made up of five clips. Each clip depicted a different character (e.g., a clown, a fairy) performing a simple action (e.g., eating, jumping). The full list of video clips is given in Appendix A. The task for the child was to describe the video to E1 who was sitting on the other side of the television such that she could not see the screen. It was made clear to the child that E1 could not see. For each video clip, E1 first simply asked the child what was happening in the video. If this did not elicit a response, then E1 asked, "What can you see?." If, in response to one of these questions, the child did not respond with a full noun and a verb, then E1 followed up with a relevant clarification question. If the child used a pronoun or null argument and a verb to describe the video (i.e., was uninformative with regard to the referent), then E1 asked "Who's verbing?" If the child used a full noun alone (i.e., was uninformative with regard to the action), then E1 asked "What is X doing?." If at any point the child did not respond at all, then E1 moved on to the next video clip.

Transcription and Coding

Transcriptions were made of all the test sessions and also for the training sessions in the child-is-speaker condition. Both E1 and E2 transcribed the child's verbal responses and pointing gestures as they occurred. Because video-recording was not permitted for ethical reasons, audio-recordings were made and checked against the online transcripts. Any discrepancies between E1 and E2's transcriptions that would have yielded different coding categories were checked and resolved by a third transcriber (the first author) from the audio-recording. Coding of pointing gestures could obviously not be checked on the audio-recording. This was not considered problematic because we only coded pointing for responses where this (plus accompanying demonstrative pronouns) was the only manner of requesting a sticker,

and in these cases, the pointing gesture was highly salient for the experimenters.

For the pre- and posttests and the training sessions in the speaker condition, the first complete attempt a child made at requesting a sticker from E1 was coded. If the child made a subsequent attempt to correct E1's choice of sticker, this final attempt was also coded separately. Reponses consisting only of pointing gestures and, potentially, accompanying demonstrative pronouns (e.g., "that one") were coded as *pointing*. Responses referring only to the character involved without the use of any disambiguating predicate (e.g., "the little girl" or "the little girl one") were coded as *naming* (even if the child also pointed to the sticker board). Responses that uniquely identified the referent (e.g., "the little girl eating" or, in the case of pictures with two characters, "the little boy stroking the dog" or "the little boy and the dog") were coded as *uniquely identifying*. Unique identifying responses were further coded for syntactic structure as containing a *relative clause* (e.g., The little girl eating) or as *other*. Rare cases where children accompanied a sentence like "the little girl going like this" with an appropriate action gesture were also counted at uniquely identifying. On occasion, some children confused "the girl" with "the mummy" or "the boy" with "the daddy" (e.g., saying "I need the girl eating" when the sticker was of the mummy eating). We accepted these responses as correct and E2 reminded children of the characters once the sticker had been selected.

For the transfer task, children's first responses were coded separately from any responses to clarification requests. All responses were coded for construction type. This provided a measure of (a) whether the child used an informative referring expression (i.e., a full noun as opposed to a pronoun or null reference) and (b) whether the children used a verb to describe the event or not. Utterances were classified into one of the six following categories (no responses fitted more than one category): *noun – verb* ("the clown is jumping" and rare anaphoric cases like "There's a clown. He's jumping"), *pronoun – verb* ("he's jumping"), *verb* ("jumping"), *noun* ("the clown"), or *no response/other* (this final category was almost entirely accounted for by children not responding at all but also included a handful of incomprehensible responses).

Utterances were coded by the first author. Twenty-four subjects (two in each condition for each age group) were randomly selected and coded by a second coder. There was very high agreement for the coding of the pre- and posttests (98% agreement, Cohen's κ = .97) and for the transfer task (98% agreement, Cohen's κ = .96).

Results

Before training, the 2- and 3-year-olds mainly pointed at or named the sticker they wanted, whereas the 4-year-olds also sometimes uniquely identified it. The younger children showed very little awareness of the inefficacy of their pointing gestures: In only 15% of cases did 2- and 3-year-olds repair a pointing gesture (by providing a naming response) after the experimenter first chose an incorrect sticker. In all other cases, they either accepted the incorrect sticker or, more often, continued pointing, upon which the experimenter chose the correct sticker. Given that the rate of pointing repairs was low, did not change substantially from pre- to posttest, and did not yield any interesting differences across conditions, we do not discuss them further here.

After training, all age groups were generally more likely to uniquely identify the sticker they needed and less likely to only point at it. The children generally used relative clauses (e.g., The daddy eating cake, The mummy softing the doggy) to uniquely identify referents (74% of the 2- and 3-year-olds' and 84% of the 4-year-olds' responses took this form). Other popular structures included conjunctions (The mummy and the dog), prepositional phrases (The mummy with the microphone), or simple declaratives (Father Christmas is sad).

The effect of training is by far most pronounced in the speaker condition as can be seen in Figures 3a to 3c, which chart the change in response types for each training condition at 2, 3, and 4 years of age, respectively. To test whether the proportions of pointing and naming responses decreased and the proportion of uniquely identifying responses increased from pre- to posttest in each of the conditions, we conducted three separate 2 (test: pre/post) × 4 (training condition) × 3 (age) analyses of variance (ANOVAs), with the proportion of each of the three major response types (pointing, naming, and uniquely identifying) as dependent variables.

Pointing

The ANOVA on pointing responses revealed a significant interaction between test and condition, $F(3, 212) = 5.08$, $p = .002$, $\eta^2 = .06$, a main effect of test, $F(1, 212) = 82.94$, $p < .001$, $\eta^2 = 0.28$, and a main effect of condition, $F(3, 212) = 3.73$, $p = .012$, $\eta^2 = .05$. These results reflect the fact that children at all ages pointed significantly less after training and that this decrease in pointing differed significantly across conditions.

Pair-wise comparisons of conditions at posttest (collapsing across age) showed that the children in
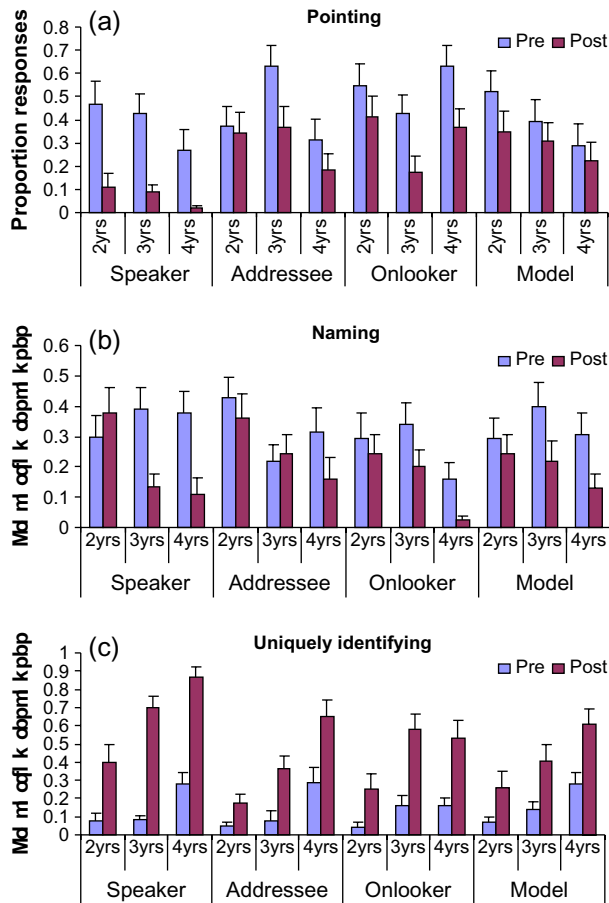
Figure 3. (a) Pointing responses as a function of training condition and test time. (b) Naming responses as a function of training condition and test time. (c) Uniquely identifying responses as a function of training condition and test time. (Error bars represent standard errors of the mean.)

the speaker condition were significantly less likely to point than the children in the three other conditions (all $ps < .001$). Thus, although children in all conditions were significantly less likely to point at posttest than at pretest training (all $ps \leq .01$), the speaker condition was most effective in reducing reliance on pointing gestures alone. These results must be understood in the context of an observed difference in rates of pointing at pretest. Pair-wise comparisons of conditions at pretest (collapsing across age) revealed that performance in all conditions was equivalent except that the children in the onlooker condition were significantly more likely to point at pretest than the children in the speaker condition ($p = .05$). But controlling for this, the percentage decrease in pointing from pre- to posttest was 32% in the speaker condition, 22% in the onlooker condition, 14% in the addressee condition, and 10% in the model description condition.

Naming

The ANOVA on naming responses revealed a significant interaction between test and age, $F(2, 212) = 4.95, p = .008, \eta^2 = .04$, a main effect of test, $F(1,212) = 28.73, p < .001, \eta^2 = .11$, and a main effect of age, $F(2, 212) = 4.67, p = .01, \eta^2 = .04$. There was no effect of condition. This reflects the fact that the 3- and 4-year-olds became less likely to simply name the sticker after training (both $ps \leq .01$) but the 2-year-olds did not. Pair-wise comparisons collapsing across condition revealed that at pretest the three age groups did not differ in the extent to which they named the sticker they wanted. However, at posttest, all the age groups differed significantly (all $ps < .02$), reflecting the fact that naming responses decreased to a greater degree as age increased. The lack of effect or interaction with condition suggests that this decrease in naming was not sensitive to different types of training.

Uniquely Identifying

The ANOVA on responses that uniquely identified referents revealed a significant interaction between test and condition, $F(3, 212) = 8.80, p < .001, \eta^2 = .11$, a significant interaction between test and age, $F(2, 212) = 10.21, p < .001, \eta^2 = .08$, a main effect of test, $F(1, 212) = 283.42, p < .001, \eta^2 = .57$, a main effect of condition, $F(3, 212) = 4.07, p = .008, \eta^2 = .05$, and a main effect of age, $F(2, 212) = 27.79, p < .001, \eta^2 = .2$. These results reflect the fact that, although the younger children were generally less likely to uniquely identify the sticker they needed, all the children were more likely to use this strategy after training and this increase differed across training conditions.

The differing effects of training condition are revealed when we compare the percentage increase in uniquely identifying responses from pre- to posttest for each condition. This increase was 41% in the speaker condition, 34% in the onlooker condition, 26% in the addressee condition, and 26% in the model description condition. Pair-wise comparisons collapsing across age revealed that all conditions were equal at pretest but at posttest the speaker condition differed significantly from all other conditions (all $ps \leq .001$), reflecting the greater increase in uniquely identifying responses in this condition.

Looking further into the interaction between age and test time, pair-wise comparisons collapsing across conditions revealed that the 2- and 3-year olds were equally likely to uniquely identify stickers at pretest but the 3-year-olds were significantly more likely to do so at posttest ($p < .001$). Thus, training had

a greater effect on 3-year-olds than on 2-year-olds. Taken in conjunction with the effect of age on naming responses, this suggests that the 2-year-olds were simply less able to produce the multiword utterances required to uniquely identify stickers. In line with this logic that older children are better able to produce longer utterances, the 4-year-olds were more likely to uniquely identify stickers than the other children at both pretest and posttest (all $ps \leq .01$).

Syntactic Structures Used to Uniquely Identify Referents

To assess the impact of hearing models on children's ability to uniquely identify stickers, we analyzed the extent to which children in each of the conditions used the same syntactic structure in their posttest responses as they had heard in training. In the addressee, onlooker, and model description conditions, the children heard each sticker described with a reduced relative clause construction during training (e.g., The girl singing). The children in the speaker condition only heard such models if they gave naming responses during training. If children were learning to use this precise syntactic construction from the models, then the children in the speaker condition might be less likely to uniquely identify referents using this precise syntactic structure than the children in the other conditions, who had heard more examples of relative clauses. To assess this, we calculated the proportion of children's unique identifying responses at posttest that contained relative clauses and compared these for each age group using Kruskal–Wallis tests (because the data were not normally distributed). There was an effect of condition in the 3-year-old group only, $\chi^2(3, N = 66) = 7.813, p = 0.05$. Mann–Whitney tests revealed that the 3-year-olds in the addressee condition were significantly *less* likely than the children in the speaker and onlooker conditions to use a relative clause in their response ($Z = 2.562, p = .01$, and $Z = 2.262, p = .024$, respectively). There is thus no suggestion that hearing more models of relative clauses led to a greater use of this structure.

Qualitative Analysis of Individual Performance

The following analyses consider the individual profiles of the children and how these changed with training. Given that each child responded six times at pretest and six times at posttest, the description of any child's response style and its change with training is potentially extremely complex. We therefore categorized children according to their predominant response style. If, at any test, four of the six responses were of one type, the child was classified by that

response type (e.g., "point"). If this criterion was not met but five of the six responses were of two types, the child was classified by those two types (e.g., "point and name"). If this criterion was not met, the child was classified as one of the three remaining combinations: "other, point, and name," "point, name, and unique," or "all response types."

Figure 4 presents the change from pre- to posttest in the number of children who were predominantly pointing, naming, or uniquely identifying stickers. Table 1 further collapses across predominant response categories to yield three qualitative categories according to the child's most informative response type at pretest and posttest (where other/pointing, naming, and uniquely identifying are considered progressively more informative). It then categorizes children according to the change in their most informative response type (e.g., "point > name" indicates that before training the child was coded qualitatively as pointing or in the "other" category (either "point," "other," or "point and other") but after training they were coded qualitatively as showing signs of naming (either "name," "point and name," or "other, point, and name"). It is worth noting that 16 children in the speaker condition began by only pointing at pretest but were able to uniquely identifying referents at posttest.

To further assess how many children improved their referring strategy with training, we classified those children who switched from only pointing or naming to using a more informative referring strategy (be it naming or uniquely identifying) as having switched to a *better* strategy. Those children whose most informative strategy at pre- and posttest was the same or had become worse with training (as was the case for only eight children) were classified as being the *same/worse*. The number of children in each category is charted for each training condition in Figure 5.
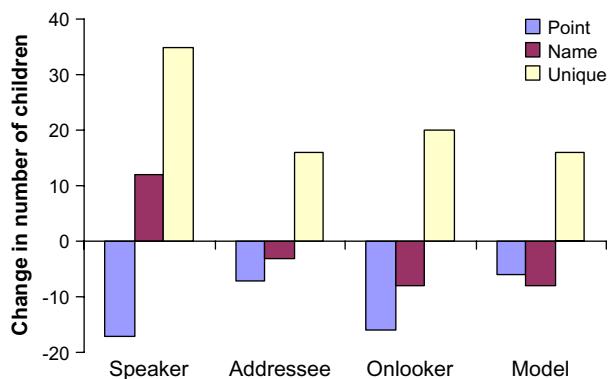

Figure 4. Change in number of children pointing, naming, and uniquely identifying stickers as a function of training condition.

Table 1
Number of Children Making Qualitative Changes in Response Style From Pre- to Posttest

Age (years)	Training	Point > point	Point > name	Point > unique	Name > point	Name > name	Name > unique	Unique > name	Unique > unique
2	Speaker	2	2	4	0	3	4	1	1
	Addressee	5	2	0	2	3	4	0	2
	Onlooker	6	3	1	0	1	5	0	1
	Model	6	1	0	0	2	4	1	3
3	Speaker	1	0	7	0	0	8	0	4
	Addressee	8	0	5	0	1	1	1	4
	Onlooker	3	1	5	0	1	5	0	5
	Model	5	0	2	1	3	5	0	4
4	Speaker	0	0	5	0	1	5	0	8
	Addressee	2	0	2	0	1	4	1	8
	Onlooker	7	0	4	0	0	2	0	7
	Model	2	1	3	1	0	3	0	8

A 4 × 2 chi-square test analyzing the results presented in Figure 5 showed that the number of children in each condition that switched to a better strategy differed significantly according to condition, $\chi^2(3, N = 224) = 13.0$, $p < .01$. Inspection of the standard residuals for each condition revealed that only the speaker condition was a major contributor to the effect.

Learning Time Course in Speaker Condition

Given the considerable advances made by the children in this training study, one might wonder how necessary having 3 days of training was and whether the children had made most of their improvements on the first day. We are able to address this question to some degree by considering the progress of the children in the speaker training
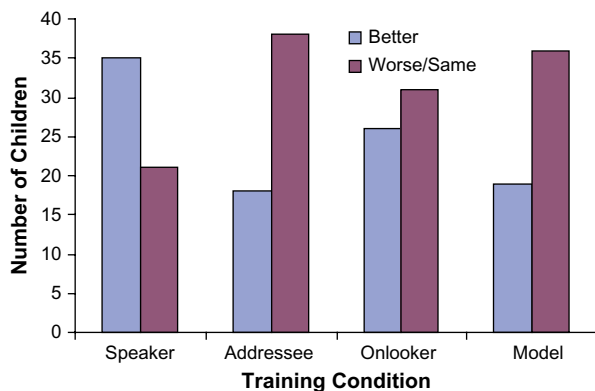

Figure 5. Number of children who did and did not switch to a more informative referring strategy after training as a function of training condition.

condition. Because these children requested stickers in training as well as in pre- and posttest, we are able to track how these requests changed over time. Figures 6a to 6c chart the children's pointing, naming, and uniquely identifying responses, respectively, from pretest, through four trainings sessions to posttest. They reveal a gradual decrease in pointing responses with time and a gradual increase in unique identifying. Naming responses only decreased for the older children.

Transfer Task

After taking the posttest, all children participated in a brief transfer task in which they needed to describe a series of five simple video clips to someone who could not see them. From a prior study (Matthews et al., 2006), we know that children aged between 2 and 4 years of age tend to describe simple scenes, for example, of a clown jumping up and down, by saying things like "he's jumping," which of course is not highly informative for an interlocutor who cannot see who "he" is. We wanted to see if training on the sticker request task would make children less likely to use pronouns in this infelicitous manner.

As an index of children's ability to use an appropriate referring expression given the addressee's perspective, we calculated the percentage of video descriptions that had a full noun argument as opposed to a pronoun. A 3 (age) × 4 (training condition) ANOVA with percentage arguments that were full nouns as the dependent variable revealed a significant effect of age, $F(2, 173) = 13.30$, $p < .001$, $\eta^2 = .133$, such that older children were more informative, but no effect of condition and no significant interaction

(a)   Pointing

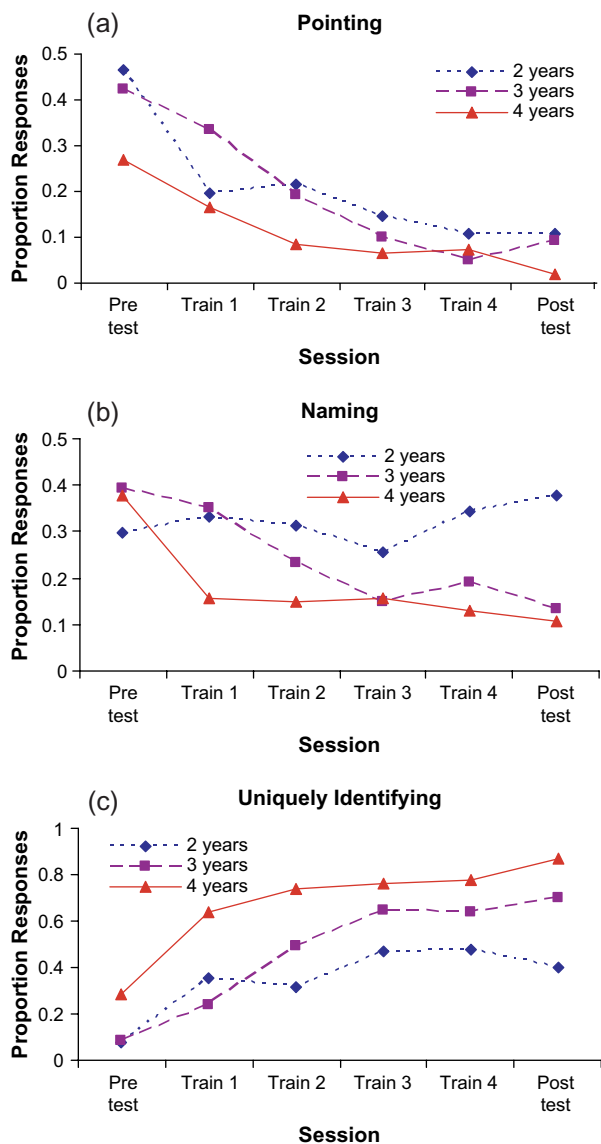
(b)   Naming

(c)   Uniquely Identifying

Figure 6. (a) Proportion of responses where children pointed as a function of time and age. (b) Proportion of responses where children named the referent only as a function of time and age. (c) Proportion of responses that uniquely identified the referent as a function of time and age.

between these two factors. Thus, either the training had no effect on children's performance in the transfer task or it had an equal effect for all conditions. We had also intended to assess whether children's ability to repair uninformative descriptions in response to feedback differed according to training condition. However, too few responses were uninformative in the first place and there were consequently insufficient repairs to analyze. An analysis of order effects did, however, suggest that the youngest children learnt from feedback during the transfer task itself.

The 2-year-olds were significantly more likely to use informative referring expressions as opposed to pronouns at the end of the test than at the beginning, $df = 1$, $\chi^2(1, N = 52) = 10.7$, $p < .001$.

As mentioned above, the fact that the children in all conditions performed equally on the transfer task raises the question as to whether training of all types had an equal, positive effect for all conditions. That is, did the children in all the conditions perform better on the transfer task than they would have done without training of any kind? To test this, we compared the responses from the current study, collapsing across training conditions, to those of the study by Matthews et al. (2006) that used precisely the same stimuli and tested children of comparable ages but had no training prior to test of any sort (current study mean ages: 2.7, 3.6, 4.5; previous study mean ages: 2.6, 3.5, 4.6). The only difference between the two studies was that the children in the current study received clarification requests if their response did not provide information about both the character and its action, whereas the children in the previous study did not. Given that this feedback had a positive effect on 2-year-olds' responses, we only compare the responses to the very first clip of each study (i.e., before feedback was given in the current study). Chi-square tests revealed that the 2- and 3-year-olds' responses did not differ significantly in the two studies. In contrast, the 4-year-olds were significantly more informative in the current study than in the study where they had received no prior training, $df = 1$, $\chi^2(1, N = 99) = 6.14$, $p < .025$. It would thus appear that only the oldest children in the current study were able to generalize from the training they had received and draw on this to become spontaneously more informative in a completely different communicative task.

Discussion

The first question this training study was designed to assess was whether young children would benefit from receiving relatively naturalistic feedback about ineffective attempts at reference. The results clearly show that they can. Even the youngest children trained in the speaker condition began to use considerably more effective referring strategies after only four 10-min training sessions spread over 3 days. Indeed, the extent to which the children stopped simply pointing to the stickers they wanted and began to describe them verbally did not differ across the age groups. Only the capacity to produce multiword utterances that uniquely described the referents was seen to develop with age such that 3- and 4-year-olds

made relatively greater gains in uniquely identifying stickers than their 2-year-old counterparts.

The second question we wanted to assess was whether the perspective from which one experiences referential repair affects learning. Again the answer appears affirmative. Although children in all three feedback conditions benefited from training, those who received feedback about their own communicative attempts (speaker condition) made by far the greatest gains. This effect of training condition was never observed to interact with age, demonstrating that the same learning experiences that were most valuable for 4-year-olds were also the most valuable for the 2-year-olds. Looking at the percentage decrease in pointing responses and percentage increase in responses that uniquely identify stickers, it would appear that the second most effective condition was the onlooker condition, in which the child observed communicative breakdown and repair from a third-party perspective. Indeed, the onlooker condition did not differ significantly from the speaker condition in our analysis of pointing difference scores. The addressee and model description conditions were also effective but always significantly less so than the speaker condition.

These results put us in a position to assess whether experiencing referential repair is a precondition of development. The fact that children improved their referring strategies in the model description condition demonstrates that observing how adults would refer to stickers (thereby strengthening relevant vocabulary and grammatical constructions) can improve referring strategies in the absence of any experience of referential repairs. This is perhaps not so surprising given that the youngest age group in this study would rarely spontaneously produce the types of syntactic structures, such as reduced relative clauses, that were most useful for uniquely identifying referents and that were heard in the model description condition (Diessel & Tomasello, 2000). Still, the gains made in this condition were markedly lower than those made when children received feedback about their own requests.

The discrepancy between the speaker and the model description conditions is all the more interesting because the speaker condition was the only condition in which children potentially did not hear any model utterances at all. Recall that, if the children in the speaker condition only pointed at the sticker they wanted, they were told that E1 could not tell where they were pointing and were asked to tell her which sticker they needed. This feedback does not provide a model referring expression. In contrast, in the three other conditions, the children heard a perfect model description for each and every sticker in training (albeit as a repair for items in the addressee and onlooker conditions where E1's first request was ambiguous). Yet, despite having far fewer examples of how to request a sticker (2-, 3-, and 4-year-olds heard models in the form of clarification requests after naming responses given during speaker condition training in only 31%, 22%, and 14% of cases, respectively), the children in the speaker condition made the greatest improvements in referring strategy. Indeed, most of the children who had begun by responding exclusively with pointing gestures (in response to which they received no model descriptions at all) made the transition to uniquely describing stickers at posttest. An interesting question for future research would be to investigate whether it was necessary to pass through a naming stage in which models, in the form of clarification requests, were heard or whether it would suffice to hear more generic clarification requests like "which one?."

So why was the speaker condition so effective? One idea is that pointing is a prepotent referring strategy that children cannot resist so long as it is available (Carlson, Moses, & Hix, 1998; Hala & Russell, 2001). It is worth noting that pilot studies established that children of this age do not point in this task if a barrier blocks the addressee's vision of the pointing gesture (see also Crain & Thornton, 1998, for techniques in eliciting relative clauses by having children describe objects for a blindfolded addressee). Furthermore, pointing may not even be that defective in an environment where an eager parent will always keep guessing at the intended meaning of a pointing gesture. Still, at some stage children will encounter less patient and competent interlocutors. The best way of helping children to develop more effective referring strategies, then, is apparently to repeatedly allow them to produce the prepotent response, have them observe its inefficiency, and encourage them to supplement a lone pointing gesture with linguistic reference. Once children break into linguistic reference, clarification requests ("Do you want the little girl eating or the little girl crying?") provide more informative models and indicate that the form of these models is to be preferred. This explanation fits well with findings in the second-language learning literature, where graduates learn to avoid language transfer errors (from their native language to their second language) and second language overgeneralizations better if they first make these errors and have them corrected than if they are merely told about the potential errors and encouraged to avoid making them (Herron & Tomasello, 1988; Tomasello & Herron, 1989).

It might be considered surprising that the addressee condition, in which children needed to respond to an experimenter's requests for stickers, was not more effective. Studies in the adult psycholinguistic literature suggest that addressees pay more attention to linguistic interactions than do onlookers and thereby derive deeper representations of dialogic interactions (Branigan, Pickering, & Cleland, 2006). Given these findings, one might expect children to benefit more from training in the addressee role than in the onlooker role and the model description condition. Yet, if anything, the onlooker condition was more beneficial. We would argue that this was the case due to the considerable difficulty children had in resolving communicative breakdown in the addressee condition.

It was very rare indeed for children to ask for clarification when given an ambiguous description of the sticker they needed. Instead, they were either confused by or oblivious to ambiguities. Thus, if E1 had only pointed at the sticker board, the children tended to look back at her confused, at which point E1 apologized and gave an unambiguous description of what she wanted. If E1 had named the sticker ambiguously, the children simply tended to guess and then E1 would correct their guess with an unambiguous description. This presumably prevented the type of learning where receiving uninformative descriptions as an addressee leads to an inference about the need for better referring strategies when speaking. Our findings here should be contrasted with those obtained with older children. Whitehurst (1976) trained 6-year-olds in the addressee role in a referential communication task and then tested them in the speaker role. Children who, as an addressee, had heard uniquely identifying descriptions went on to be significantly more informative speakers than children in a control condition who had received no training. Surprisingly, though, children who had heard ambiguous descriptions as addressees, went on to perform significantly worse than the control group—that is, they were actively attempting to be ambiguous in their descriptions. Clearly, 6-year-olds can learn a lot about styles of communication from being the recipient of both sufficient and insufficient information.

In the onlooker condition, the children were able to observe an ineffective attempt at reference, an identification of the problem by the addressee, a request for clarification, and finally unique identification of the sticker. This process could be more effective simply due to the sheer number of model descriptions the child heard (both in the clarification request and in the final unambiguous reference). However, in light of the results in the speaker condition, we consider it

more likely that it was a combination of these models and of a less perplexing interaction that led to the relatively large advances observed in the onlooker condition (see also Keysar & Henly, 2002, for a discussion of the privileges of the onlooker viewpoint).

Over and above explanations of the benefits of different training conditions, we must also ask what it is precisely that children learnt to do in this study. We have seen from the results of the model description condition that at the very least, the children had learnt to use linguistic structures that they are not normally inclined to depend on. The question remains as to how much they learnt about the necessity of these descriptions given the communicative context. Caution is needed here because previous training studies have observed children to learn that longer utterances are better for referential communication tasks without learning that the reason why longer is better is because longer utterances are generally more informative. Thus, Whitehurst et al. (1981) observed that 5-year-olds performed better on referential communication tasks after hearing models of long messages that did not uniquely identify referents (e.g., "the big red one" in a context where there was more than one big red object) than after hearing models of short messages that did (e.g., "the green one" in a context where there was only one green object).

Were the children in the current training study simply learning to produce longer utterances without understanding why they were needed? The success of the children in the current speaker condition who did not hear any models suggests that the learning we observed was not of the completely mindless "the name of this game is just to produce a lot of words" type. These children had to learn that pointing was not an effective strategy for obtaining what they needed. Only once they themselves switched to linguistic strategies were language models provided in clarification requests, which the children could subsequently learn from in order to produce adequately informative linguistic descriptions.

Whether the children fully understood that longer linguistics descriptions were adequate only by virtue of being sufficiently informative cannot be fully assessed from the current study. One interesting follow-up would be to include arrays of differing complexity at posttest, some of which require complex descriptions to uniquely identify referents and some of which would only require a name (e.g., if there was only a sticker of a boy and a girl on the board, referring to the girl as "the little girl eating an ice-cream" would be redundant). If children modulated the length and informativity of their referring expressions according to the complexity of the array, then we would have

good reason to believe that they had understood the communicative function of complex referring expressions. Unfortunately, though, children might well become redundant in this task in much the same way adults do (Brennan & Clark, 1996) and so whereas finding differentiation of message length as a function of array complexity would suggest insight into communication, a finding of redundancy would not necessarily reflect a lack of such insight.

The issue of array complexity bears further discussion in relation to what children learnt in the current study. One component of successful reference is the identification of potential for misunderstanding based on what is in common ground or perceptually salient for others. This process is a challenge even for well-educated university students who have been observed to forget when and with whom more explicit referring expressions are called for (Horton & Gerrig, 2002, 2005) or, conversely, to become hyperexplicit when there are numerous potential referents in the perceptual context or prior discourse (Arnold & Griffin, in press; Haywood, Pickering, & Branigan, 2005). Similarly, as referential arrays become more complex, children's referring strategies to the same object may switch from linguistic reference to pointing (c.f. studies 1 and 3 in Pechmann & Deutsch, 1982) suggesting that too much complexity makes children give up on language and get things started with gesture. It is most probable that, at the beginning of this study, children were unaware that the number and similarity of the stickers on the board could cause communication difficulty. Thus, it is plausible that the training we provided not only developed knowledge of referring expressions but also heightened the children's appreciation of their addressee's perceptual environment. These observations contribute to the ongoing discussion of the contribution of language to the development of social cognitive skills (Astington & Baird, 2005). They also give rise to a theory of communication whereby one's background goal is generally to be an informative communicator but one's ability to be so is not tightly constrained at every conversational turn.

Finally, we might ask how transferable the knowledge that the children developed in the current study is to other communicative settings. The only evidence we have that the children went beyond learning to adapt to the contextual constraints of the current study and gained insight into the nature of communication more broadly comes from the transfer task. By comparing the children's responses in this test with those of a prior study, we found that the 4-year-olds that had participated in the current study (in whichever training condition they had been assigned

to) were more informative in describing videos than were children who had received no prior training of any sort. This somewhat limited observation of transfer is not unusual in referential training studies (Asher & Wigfield, 1981). It is also fair to say that our transfer task was very different in nature to the training task in that (a) it involved describing rather than requesting and (b) the referents were out of the addressee's sight rather than in a dense but perceptually available array. Transferred improvement on such a different task was clearly too great a leap to make for the younger participants and reinforces the observation that even when children have the necessary skills for successful reference, they do not always put them together (Whitehurst & Sonnenschein, 1981).

Of course, even adults often do not make the most of their abilities to communicate effectively (Epley, Morewedge, & Keysar, 2004; Keysar & Henly, 2002; Keysar, Lin, & Barr, 2003) and we would not want to set the learning bar artificially high. To adequately describe development and its endpoints, then, we will need to establish how mature speakers communicate, where they are vulnerable to misunderstanding, and how they tend to resolve difficulties. Recent studies of adult abilities to convey information to others have revealed not only adept choice of referring expressions but also an ability to set up entire referential episodes, where the speaker guides their addressee to incrementally formulate representations of referents and the scenes to which they belong (Smith, Noda, Andrews, & Jucker, 2005). Clearly, adult speakers are exquisitely sensitive to the goals of communicative exchange even if their execution of these goals can at times leave a lot to be desired. The results of this training study suggest that young talkers have the same requisite desire to be understood and that their increasing experience of dialogue is central to developing the tools for meeting their aims.

References

Anselmi, D., Tomasello, M., & Acunzo, M. (1986). Young children's responses to neutral and specific contingent queries. *Journal of Child Language*, *13*(1), 135–144.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, *24*, 65–87.

Arnold, J., & Griffin, Z. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory & Language*, *56*, 521–536.

Asher, S. R., & Wigfield, A. (1981). Training referential communication skills. In W. P. Dickson (Ed.), *Children's oral communication skills* (pp. 105–126). New York: Academic Press.

Astington, J., & Baird, J. (Eds.). (2005). *Why language matters for theory of mind*. Oxford, UK: Oxford University Press.

Branigan, H., Pickering, M., & Cleland, A. (2006). Syntactic alignment and participant role in dialogue. *Cognition, 104,* 163–197.

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.

Carlson, S., Moses, L., & Hix, H. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development, 69,* 672–691.

Crain, S., & Thornton, R. (1998). *Investigations in universal grammar: A guide to experiments on the acquisition of syntax and semantics.* Cambridge, MA: MIT Press.

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

Dickson, W. P. (1982). Two decades of referential communication research: A review and meta-analysis. In C. J. Brainerd & M. Pressley (Eds.), *Verbal processes in children.* New York: Springer.

Diessel, H., & Tomasello, M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics, 11,* 131–151.

Epley, N., Morewedge, C., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology, 40,* 760–768.

Gallagher, T. (1977). Revision behaviors in the speech of normal children developing language. *Journal of Speech and Hearing Research, 20,* 303–318.

Glucksberg, S., Krauss, R. M., & Weisberg, R. (1966). Referential communication in nursery school children: Method and some preliminary findings. *Journal of Experimental Child Psychology, 3,* 333–342.

Golinkoff, R. M. (1986). 'I beg your pardon?': The preverbal negotiation of failed messages. *Journal of Child Language, 13,* 455–476.

Golinkoff, R. M. (1993). When is communication a 'meeting of minds'? *Journal of Child Language, 20,* 199–208.

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69,* 274–307.

Hala, S., & Russell, J. (2001). Executive control within strategic deception: A window on early cognitive development? *Journal of Experimental Child Psychology, 80,* 112–141.

Haywood, S., Pickering, M., & Branigan, H. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science, 16,* 362–366.

Herron, C., & Tomasello, M. (1988). Learning grammatical structures in a foreign language: Modelling versus feedback. *French Review, 61,* 910–923.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory & Language, 47,* 589–606.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition, 96,* 127–142.

Keysar, B., & Henly, A. (2002). Speakers' overestimation of their effectiveness. *Psychological Science, 13,* 207–212.

Keysar, B., Lin, S., & Barr, D. (2003). Limits on theory of mind use in adults. *Cognition, 89,* 25–41.

Krauss, R. M., & Glucksberg, S. (1969). The development of communication competence as a function of age. *Child Development, 40,* 255–266.

Lefebvre-Pinard, M., & Reid, L. (1980). A comparison of three methods of training communication skills: Social conflict, modelling and conflict modelling. *Child Development, 51,* 179–187.

Liszkowski, U. (2006). Infant pointing at twelve months: Communicative goals, motives and social-cognitive abilities. In N. Enfield & S. Levinson (Eds.), *The roots of human sociality: Culture, cognition and interaction* (pp. 153–178). Oxford, UK: Berg.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Reference and attitude in infant pointing. *Journal of Child Language, 34,* 1–20.

Lloyd, P., Mann, S., & Peers, I. (1998). The growth of speaker and listener skills from five to eleven years. *First Language, 18*(52), 81–103.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk (Vol. 2: The database).* London: Lawrence Erlbaum.

Matthews, D. E., Theakston, A., Lieven, E., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics, 27,* 403–422.

O'Neill, D., & Topolovec, J. (2001). Two-year-old children's sensitivity to the referential (in)efficacy of their own pointing gestures. *Journal of Child Language, 28,* 1–28.

Pechmann, T., & Deutsch, W. (1982). The development of verbal and nonverbal devices for reference. *Journal of Experimental Child Psychology, 34,* 330–341.

Robinson, E. J., & Robinson, W. P. (1981). Ways of reacting to communication failure in relation to the development of children's understanding about verbal communication. *European Journal of Social Psychology, 11,* 189–208.

Robinson, E. J., & Robinson, W. P. (1985). Teaching children about verbal referential communication. *International Journal of Behavioral Development, 8,* 285–299.

Smith, S., Noda, H., Andrews, S., & Jucker, A. (2005). Setting the stage: How speakers prepare listeners for the introduction of referents in dialogues and monologues. *Journal of Pragmatics, 37,* 1865–1895.

Sonnenschein, S., & Whitehurst, G. (1984). Developing referential communication: A hierarchy of skills. *Child Development, 55,* 1936–1945.

Tomasello, M., Anselmi, D., & Farrar, M. J. (1984/1985). Young children's coordination of gestural and linguistic reference. *First Language, 5,* 199–210.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78,* 705–722.

Tomasello, M., & Herron, C. (1989). Feedback for language transfer errors: The garden path technique. *Studies in Second Language Acquisition, 11,* 385–395.

Whitehurst, G. (1976). The development of communication: Changes with age and modeling. *Child Development*, *47*, 473–482.

Whitehurst, G., & Sonnenschein, S. (1981). The development of informative messages in referential communication: Knowing when versus knowing how. In W. P. Dickson (Ed.), *Children's oral communication skills* (pp. 127–141). New York: Academic Press.

Whitehurst, G., Sonnenschein, S., & Ianfolla, B. (1981). Learning to communicate from models: Children confuse length with information. *Child Development*, *52*, 507–513.

Appendix A

Description of Materials Used: Target Stickers and Video Clips

The Bumbles Have a Party: Dad dancing, girl eating cake, clown crying, boy stroking dog, mum singing, dog chasing clown. *The Bumbles at Christmas*: Dad eating cake, dog chasing Father Christmas, boy dancing, girl singing, mum tickling dog, Father Christmas crying.

The Bumbles Go to the Zoo: Dad drawing, girl feeding monkey, boy crying, mum eating bananas, elephant sleeping, giraffe licking mum. *The Bumbles Go to the Farm*: Sheep jumping, dad eating carrots, girl crying, mum sleeping, pig chasing dad, boy driving tractor.

The Bumbles Go on Holiday: Dad eating ice-cream, boy riding donkey, mum reading, dog sleeping, girl swimming, donkey chasing dad. *The Bumbles Stay at Home*: Dad reading, mum washing dog, girl drawing, cat sleeping, boy eating, dog licking mum. *Transfer task video clips*: Fairy eating, clown crying, witch jumping, Father Christmas fell over, fairy crying.