Two- and Four-Year-Olds Learn to Adapt Referring Expressions to Context: Effects of Distracters and Feedback on Referential Communication

Danielle Matthews,\textsuperscript{a} Jessica Butcher,\textsuperscript{b} Elena Lieven,\textsuperscript{c} Michael Tomasello\textsuperscript{d}

\textsuperscript{a}Department of Psychology, University of Sheffield
\textsuperscript{b}Max Planck Child Study Centre, University of Manchester
\textsuperscript{c}Max Planck Child Study Centre, University of Manchester and Max Planck Institute for Evolutionary Anthropology
\textsuperscript{d}Max Planck Institute for Evolutionary Anthropology

Received 18 February 2010; received in revised form 14 April 2011; accepted 29 April 2011

Abstract

Children often refer to things ambiguously but learn not to from responding to clarification requests. We review and explore this learning process here. In Study 1, eighty-four 2- and 4-year-olds were tested for their ability to request stickers from either (a) a small array with one dissimilar distracter or (b) a large array containing similar distracters. When children made ambiguous requests, they received either general feedback or specific questions about which of two options they wanted. With training, children learned to produce more complex object descriptions and did so faster in the specific feedback condition. They also tended to provide more information when requesting stickers from large arrays. In Study 2, we varied only distracter similarity during training and then varied array size in a generalization test. Children found it harder to learn in this case. In the generalization test, 4-year-olds were more likely to provide information (a) when it was needed because distracters were similar to the target and (b) when the array size was greater (regardless of need for information). We discuss how clear cues to potential ambiguity are needed for children to learn to tailor their referring expression to context and how several cues of heuristic value (e.g., more distracters > say more) can promote the efficiency of communication while language is developing. Finally, we consider whether it would be worthwhile drawing on the human learning process when developing algorithms for the production of referring expressions.

Keywords: Referring expressions; Language acquisition; Training
1. Introduction

It is broadly agreed that human communication is all about maximizing our chances of a successful exchange while minimizing cognitive effort for both listener and speaker (e.g., Bard et al., 2007; Clark & Wilkes-Gibbs, 1986; Dale & Reiter, 1995; Wilson & Sperber, 2006). Given this general consensus, the question that faces the computational linguist and the psycholinguist alike is precisely how language systems (be they artificial or real) are built to meet these requirements. In our contribution to this topic, we bring evidence from developmental psychology to bear on this question. We hope that by observing the building in action, so to speak, we will be better placed to answer this shared question.

The development of reference gets well under way in the first year of life, when infants begin to enjoy communicating with others about external objects in the world (Bruner, 1983). The most obvious way in which children achieve this is with pointing gestures that emerge around the first birthday (Tomasello, Carpenter, & Liszkowski, 2007) and with their first words, which often emerge shortly after (Fenson et al., 1994). Reference, be it gestural or vocal, is nebulous at the outset: It is impossible to say when a child first produced a real pointing gesture or word in the sense we typically intend. Nonetheless, most children go on to rapidly develop a sizeable repertoire of referring expressions and, as they do, it becomes interesting to question how they learn to choose appropriate expressions given the context and, more specifically, given the accessibility of the referent for the addressee (Ariel, 1988, 1990).

In the current study, we will present two training studies designed to assess how children learn to produce referring expressions that are sufficiently informative but not unduly redundant. Both of the studies are experimental. We should note, however, that there is a rich cross-linguistic literature investigating children’s production of referring expressions in naturalistic and narrative contexts (e.g., Allen, 2007; Bamberg, 1987; Bates, 1976; Clancy, 2003; Greenfield, 1979; Hickmann & Hendriks, 1999; Küntay, 2002; Küntay & Özyürek, 2006; Rozendaal & Baker, 2008; Salazar Orvig et al., 2010; Serratrice, 2004; Skarabela, 2007). Naturalistic studies suggest that, from the onset of multiword speech (approximately 24 months), children adapt their production of referring expressions according to whether something is accessible for the listener. However, as Clancy (2003) points out, the referential form a child uses in any given utterance will depend to a large extent on its grammatical role and on the lexical properties of the utterance. Consequently, it is difficult to establish how much children’s awareness of the informational status of different referring expressions depends on the scaffolding of the restricted contexts in which they are used (cf., Karmiloff-Smith, 1981). Experiments have thus been designed to complement this rich naturalistic data.

Experimental studies so far have established that children as young as 2;3 (2 years 3 months) choose whether to gesture to the location of a hidden toy depending on whether their parent saw the hiding event (O’Neill, 1996; see also O’Neill & Topolovec, 2001). In fact, even 12-month-olds have been found to point appropriately for knowledgeable versus ignorant partners (Liszkowski, Carpenter, & Tomasello, 2008). With the advent of language, children need to move beyond the choice of whether to communicate to the choice of what linguistic form to use for communication. Looking at verbal reference, Matthews, Theakston,
Lieven, and Tomasello (2006) found that two 1⁄2-year-olds chose whether to refer to a character lexically (‘A man jumping’) or to use null reference (‘Jumping’) depending on whether it was given in the previous discourse (see also Salomo, Lieven, & Tomasello, 2009). In addition to this, 3-year-olds chose whether to produce an informative referring expression (in contrast to a pronoun or null reference) according to whether their interlocutor could see what they were talking about. Four-year-olds were better able to express inaccessible referents lexically and accessible ones pronominally (as an adult English speaker would), although they were still prone to errors (see also Campbell, Brooks, & Tomasello, 2000; Wittek & Tomasello, 2005). By 5 years, other studies demonstrate that children are more likely to modify nouns with adjectives when referring to an object with a similar distracter in common ground as opposed to privileged ground (Bahtiyar & Küntay, 2009; Davies & Katsos, 2010; Nadig & Sedivy, 2002; Nilsen & Graham, 2009).

It is important to note, though, that, in the above studies, children are deemed sensitive to an accessibility cue if they are more likely to use informative referring expressions when it is required (because a referent is inaccessible) than when it is not. This does not mean that children use the appropriate expression all of the time. For example, in the Matthews et al.’s (2006) study, on trials where a referent was inaccessible, about a third of the referring expressions produced by 4-year-olds were, inappropriately, pronouns. Thus, even though the foundations for appropriate production of referring expressions appear to be in place in infancy, it will clearly take quite some time for children to become adept in using the complete toolbox of referring expressions that any language has to offer. It is therefore not surprising that decades of research have found that children have real difficulty in selecting appropriate referring expressions right up to adolescence (see Dickson, 1982, for a review of early work in this area).

The question that interests us in this study concerns the experiences that facilitate the use of appropriate referring expressions in context. To address this, it is helpful to conduct training studies. The critical fact these studies are based on is that, when asked for clarification about their ambiguous requests (e.g., “Which blue ball?”), even young children will often provide the necessary disambiguating information (e.g., Anselmi, Tomasello, & Acunzo, 1986; Deutsch & Pechmann, 1982; Gallagher, 1977; Golinkoff, 1986; Wilcox & Webster, 1980). This ability to repair is quite striking and demonstrates that young children have a strong desire to be understood (cf., Shwe & Markman, 1997) and are far from completely oblivious to the need to be informative in conversation (see also Forrester & Cherington, 2009, for a Conversational Analysis approach to self-repair in children). It suggests that development relies on an early ability to engage communicatively with one’s interlocutor coupled with a protracted period of learning from trial and error in order to be able to convey ever more intricate messages in ever more demanding contexts.

Evidence that repairing failed requests can indeed facilitate communicative development was reported by Robinson and Robinson (1985), who found that asking 5-year-olds to clarify ambiguous descriptions on one trial led them to be more informative from the outset on subsequent trials. In a recent training study, Matthews, Lieven, and Tomasello (2007) demonstrated that the benefits of receiving feedback about one’s own communicative attempts extend to younger children and outdo any benefits derived from training with relevant
vocabulary and grammar. In this study, 2-, 3-, and 4-year-old children were pre- and posttested for their ability to request stickers from a dense array of similar alternatives. Between test sessions, children were assigned to a training condition in which they either (a) asked for 24 stickers from an adult and received feedback if their requests were ambiguous ("speaker" condition), (b) responded to 24 (sometimes ambiguous) adult requests for stickers, (c) observed one adult ask another for 24 stickers (sometimes ambiguously), or (d) heard 24 model descriptions of stickers (to train knowledge of relevant vocabulary and grammatical constructions such as reduced relative clauses). Children in all four conditions improved their referring strategies after training. The speaker condition was the most effective. In this condition, if a child requested a target by pointing at the array of stickers, she/he was asked, "Which one?" and if she/he asked for a sticker ambiguously (e.g., saying "I need the girl" when there were two stickers that met this description), then the experimenter would explicitly ask which option the child had meant (e.g., "Do you need the girl eating or the girl singing?").

In the above study, even very young children responded to clarification requests and learned from doing so. In this study, we ask precisely what they were learning and how. Ideally, what children would have learned about would be their interlocutor's needs for information (in order for them to identify a target). However, previous research would suggest that 2-year-olds might not be able to properly assess their interlocutor's need for information in such settings. Whitehurst, Sonnenschein, and Ianfolla (1981) demonstrated that even 5-year-olds confuse description length with informativity (assuming that longer utterances are more informative regardless of their content). It is quite possible, therefore, that training children to be more informative is in fact only successful insofar as it makes them more verbose. That is, children may just have learned to produce longer descriptions of stickers without any insight into why longer was, in the case at hand, better. We address this issue in the current studies by manipulating how much information is required to identify a target and measuring the degree to which children provide redundant information in their descriptions.

When it comes to assessing how children were learning to be more informative, we were interested to investigate whether some forms of feedback were better suited for learning than others. In our previous training study, children's pointing gestures met with a general request for clarification, "Which one?" whereas their ambiguous descriptions met with more specific feedback, "Do you need the girl eating or the girl singing?" It is possible that this latter style of specific feedback would highlight to children the potential ambiguity of their original response and promote learning faster. However, in doing so, it might simultaneously encourage redundancy in trials where extra information was not required. That is, because it is easier to respond to a specific feedback request, children may learn in a more shallow fashion (to produce descriptions of the form "I need the X VERBing") and have difficulty generalizing to new trials.

In Study 1, we investigated both what and how children learn during training by manipulating two factors: array size and feedback style. To assess whether children understood that longer descriptions were only necessary because they would uniquely identify a referent (given the target's similarity to alternatives), we placed the target sticker in one of two contexts (see Fig. 1): either with only one other dissimilar sticker (2-sticker array condition) or in the middle of three stickers, two of which were similar to the target on a given dimension
If children learn from training that it is better to produce longer utterances, then they should produce long descriptions regardless of array size. On the other hand, if children have some appreciation of why the information in a description is needed, they might give longer descriptions of target stickers when they are located in 4-sticker arrays as compared to 2-sticker arrays.

To further investigate why children learn more from gaining feedback about their own requests than from hearing model descriptions of referents, we had two training conditions. In the general feedback condition, the experimenter indicated only that she could not identify the target if the child’s request was ambiguous. For example, if the child asked for “The girl” from an array containing two similar girl stickers, then the experimenter would ask “Which girl?” This feedback leaves the child to think of a suitable repair. In the specific feedback condition, clarification requests contained a model description that could be used as an answer. For example, in response to “The girl,” the feedback might be, “Do you need the girl eating or the girl reading?” In this case, the child can select the correct option and repeat it. If hearing models of the type of description required (compared to a description of a distracter) is particularly helpful, then children in the specific feedback condition should show greater improvements. It is further possible that the type of feedback children receive would interact with array size. For example, children in the specific feedback condition might become more descriptive but less able to adapt their descriptions according to array size, giving lengthy descriptions regardless of whether they were necessary.

2. Study 1

2.1. Method

2.1.1. Participants

Eighty-four normally developing, monolingual, English-speaking children were included in the study (33 boys, 51 girls). There were thirty-eight 2-year-olds (mean age: 2;8, range: 2;2–3;11).
2;1–3;2) and forty-six 4-year-olds (mean age: 4;1, range: 3;8–4;7). Children were matched for gender and age in months across the between-subjects feedback conditions. A further seven children participated but were excluded because they (a) did not complete one of the testing sessions, (b) consistently asked for nontarget stickers, or (c) turned out to be seeing a speech and language therapist. Testing took place in a quiet area in the children’s day care center or at the Max Planck Child Study Centre, Manchester, UK. The children were predominantly White and middle class. Four children were from ethnic minorities.

2.1.2. Materials and design

The experiment followed a 2 (array context) × 2 (feedback condition) × 2 (age) design, with array context as a within-subjects factor and feedback condition and age as between-subjects factors.

Six picture books were made about the adventures of a family. Each book contained six unrelated pictures of family members performing simple actions. For testing, each book was printed in two versions that were identical except that one had all the characters missing (see Fig. 2). The child’s task was to make the incomplete book the same as the complete one by putting stickers in it. The stickers were obtained from Experimenter 1 (henceforth E1), who kept them on a board out of the child’s reach. Beneath the board, on the floor was a red cardboard ring. The child was asked to go to the circle before asking for a sticker. This ensured that children would come up to the board and look at the stickers before requesting one.

For 2-sticker array contexts, the target sticker was placed directly above another, completely different sticker (see Fig. 1). For 4-sticker array contexts, the target sticker was placed above a different sticker and between two similar stickers (one had the same character as the target, the other had the same action).

Fig. 2. Example of picture book pages.
Each child completed all six books in the same order and requested 36 stickers in total, 18 in each array size. Every other sticker was presented in a 4-sticker array (counterbalanced such that each target sticker was seen an equal number of times in each array size for the experiment as a whole).

2.1.3. Procedure

Before testing began, the child and E1 sat together at a table and played a “snap” game. Six pairs of identical cards were spread out on the table. E1 took one card and asked the child to find the other card that was “the same as this one.” The game continued until all six pairs of cards had been matched. This task ensured that children understood the term “the same” and would be able to make their incomplete book “the same” as E2’s book.

After the snap game, E1 went to the sticker board on the other side of the room. E2 came to the table and presented the two versions of the first picture book, explaining one was for her and the other was for the child. E2 asked him/her to point to each of the characters on the cover of the picture book, which all the children could do. 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 complete it. She explained that the child could walk over the E1, stand in the “magic red circle” and ask her for a sticker. A red circle was taped onto the floor so that the experimenters could encourage children not to ask for the sticker until they were in the magic circle. This ensured that children were in a position to see the array of stickers before they produced a request. E2 made it explicit that from where E1 was sitting she could not see their books as there was a large box on the table that blocked her view. E2 turned the cover pages of her own and the child’s book over to reveal the first picture in its complete and incomplete versions. She asked the child if she/he could go over to stand in the circle and ask E1 for the sticker that would make their books the same.

E1 encouraged the child to come and retrieve the sticker she/he needed. Once the child was in front of the board, both she/he and E1 could see all the stickers, which were high enough on the wall to make any pointing gesture on the part of the child ambiguous.

Once a child requested a sticker, the feedback she/he received differed according to (a) how informative the request was, (b) the size of the array, and (c) which condition she/he had been randomly assigned to. If the array size was 2, then the process of giving feedback was terminated as soon as the child provided enough information to identify the target sticker (e.g., “The Daddy one”). If the array size was 4, the feedback was continued until the children provided a complex description (e.g., “The Daddy eating”). Children in the specific feedback condition were given feedback that provided the child with two possible descriptions, only one of which was appropriate for their target. The children in the general feedback condition met with clarification requests that indicated more information was required about the target but did not provide or contrast specific options. Feedback was contingent on the child’s response as is now described.

In both conditions, if the child first pointed to a sticker (and/or said “that one”), then E1 first asked, “Can you tell me which one you need, I can’t tell where you are pointing?” If this failed to elicit a sufficient description, E1 asked for clarification. In the specific feedback condition, the child was asked, for example, “Do you need the dad or the boy?” In the
general feedback condition, the question was, “Who do you need?” If in response, the child pointed once again, E1 asked the child if she needed one of the nontarget stickers by saying in the specific condition, for example, “Do you need the daddy dancing?” or in the general condition “Do you need this one?” If the child rejected this sticker (shaking the head or saying “no”), E1 then asked if the child needed the target sticker, describing it fully in the specific condition (e.g., “Do you need the daddy eating carrots?”) or by asking, “Do you need this one?” in the general condition. If she/he accepted the target (nodding or saying “yes”), E1 handed over the sticker.

If at any point the child gave a simple description of the sticker (i.e., one that was informative enough in the 2-sticker array context but not in the 4-sticker array context), then, in the 2-sticker array context, E1 handed the target sticker over. In the 4-sticker array context, E1 asked for clarification in the specific condition by fully describing two potential referents, of which one was the target (e.g., “The daddy eating carrots or the daddy dancing?”). In the general condition, the child was asked, for example, “Which daddy?” If the child answered with the appropriate complex description, then E1 gave her the correct sticker. If she/he persisted in giving a simple, insufficient description or regressed to pointing, then E1 followed the procedure for pointing responses and asked whether the child needed one of the nontarget stickers.

The study continued in the above fashion until the child had requested all 36 stickers. Testing took place over two sessions (three new books made in each session) that were a maximum of 6 days apart (four 2-year-olds, two in each condition, exceeded this and had test sessions 2 weeks apart).

2.1.4. Transcription

Both E1 and E2 transcribed each child’s verbal responses and pointing gestures as they occurred. Video recordings were made in the laboratory, and audio recordings were made in day care centers. Rare discrepancies between E1 and E2’s transcriptions that would have yielded different coding categories were checked from the recordings. Transcription of pointing gestures could not be checked on the audio recording but discrepancies in these cases were very rare indeed. In such cases where one experimenter transcribed a pointing gesture but the other did not, we assumed that the gesture had occurred (this was also consistent with E1’s subsequent feedback that could be heard on the recording).

2.1.5. Coding

Each turn in a child’s responses was coded as either pointing, simple, complex, or incorrect. Responses consisting only of pointing gestures and, potentially, demonstrative pronouns (e.g., “That one”) were coded as pointing. If responses contained just enough information to identify a sticker in the 2-sticker array context, they were coded as simple. Simple responses thus either (a) named a character (e.g., “The little girl one”), (b) referred to an action or property of the character (e.g., “eating carrots”), or (c) gave partial information about the sticker’s spatial location (e.g., “Top one”). Responses that gave information sufficient to locate the target in a 4-sticker array were coded as complex (e.g., “The girl eating,” “The boy at the top in the middle”). Rare cases where children gave information
with a gesture or baby sign (e.g., ‘‘The lady’’ + snoring action for the lady sleeping or an elephant gesture/sign for a sticker where an elephant was the target character) were included and coded as simple or complex according to whether they would uniquely identify a target in the 4-sticker array case (complex) or only the character or some property of it (simple). If a child was slightly inaccurate in his or her description (e.g., confused the gender or reversed the roles of characters [e.g., saying Santa chasing a doggy instead of A doggy chasing Santa] but nonetheless uniquely identified it), these responses were included. Occasionally, children explicitly asked for an incorrect sticker or accepted a nontarget sticker when it was proposed by E1. In these cases, E1 would hand this sticker to the child and E2 would point out that it was not the right one. These responses were coded as incorrect.

Eight subjects (two in each condition of each age group) were randomly selected and coded by a second coder, the first author. We focus in this study on children’s first attempts at reference (how, on any given trial, the child first requested the sticker before the experimenter gave feedback). We calculated coding agreement for this measure and agreement was very high (99% agreement in all cases, Cohen’s $\kappa > .9$ in all cases).

2.2. Results

Children in both age groups often pointed in order to request stickers at the very beginning of the study but rapidly began to verbally describe them after a few turns with feedback. They gradually became more descriptive with time. When children provided complex descriptions, they often used reduced relative clauses (e.g., ‘‘the daddy dancing’’), prepositional phrases (‘‘the boy with the dog’’), conjunctions (e.g., ‘‘the boy and the dog’’), and ungrammatical combinations that nonetheless contained enough information to identify a target (e.g., ‘‘girl cake’’). A few children repeatedly employed less common structures (e.g., ‘‘the one where the little boy is sitting down in the tractor’’), gestured an action in place of a verb (e.g., ‘‘the girl doing this’’), or clarified first attempts by giving spatial information (e.g., ‘‘in the middle’’). By far, the most common complex description type was the reduced relative clause. Considering all the utterances the children produced over the experiment as a whole, the children in the specific feedback condition produced approximately twice as many relative clauses compared to those in the generic feedback condition. This reflects the fact that, in the specific feedback condition, children heard reduced relative clauses in the experimenter’s clarification requests. When clarifying, children could directly reuse the appropriate description. When responding on a new trial, they could give descriptions with the same syntactic structure (following the pattern ‘‘The X VERBing’’). As a consequence, children in the specific feedback condition took fewer turns to reach a sufficient description. Likewise, all children were faster to reach a sufficient description on 2-sticker array trials than on 4-sticker array trials.

The critical questions we have are whether children learned better from one type of feedback than another and whether children learned to be as descriptive as possible at all times or rather were more likely to produce complex descriptions when they were necessary (4-sticker array) than when a simple description would suffice (2-sticker array). To test this, we consider children’s first attempts at requesting a sticker (i.e., what they did before
receiving any feedback for the given trial). Fig. 3 shows the number of times children’s first attempts at requesting a sticker were complex verbal descriptions (which are necessary when the array size is 4 but overinformative when the array size is 2).

To assess whether the complexity of verbal descriptions varied according to age, time, feedback condition, and sticker array size, we fitted a mixed-effects logistic regression model to the data using Laplace approximation (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Dixon, 2008; Gelman & Hill, 2007; Jaeger, 2008). The outcome variable was whether the child’s description was complex (1) or not (0). Child ($n = 84$) was added as a random effect on the intercept. Throughout the study we explore theoretically motivated models by starting with a simple model with all fixed and random effects, adding two-way interactions in their possible combinations and checking for improvement in fit.

Fig. 3. Frequency of complex responses in Study 1 as a function of age, feedback condition, array size, and time. (Due to counterbalancing, the maximum possible value per time step = 12.)
In Table 1, we report the parameters of a model (Model 1) that contained all experimental variables and the following two-way interactions: time · age, time · feedback condition, time · array, array · feedback condition. This had a significantly better fit to the data than models that contained all experimental variables but no interactions, \( \chi^2(6) = 31.98, p < .0001 \), and was a borderline better fit than a model with all experimental variables and just the two-way interactions time · age and time · feedback condition, \( \chi^2(2) = 5.42, p < .066 \). There was no improvement in fit when we added further interactions. For Model 1, the estimated intercepts for the children varied with a standard deviation of 1.49.

The significant interaction of age with time reported in Table 1 reflects the fact that, with training, the older children became more likely to use complex expressions than the younger children. Likewise the time · feedback interaction indicates that those in the specific feedback condition became more likely to use complex expressions than those in the general feedback condition. The borderline interactions of array with time and feedback suggest that children may have been distinguishing between the arrays when deciding whether to produce a complex expression. The critical question is whether they were doing this at the end of training. If children are simply learning to be verbose, then they may adapt their referring expressions according to array size early on and then override this after training with a simple tendency to be as descriptive as possible at all times. To test whether an effect of array held at the end of the training study, when children had learned to use a greater number of complex descriptions, we analyzed children’s first attempts at reference on the final 10 trials of Study 1 (five with 4-sticker arrays and five with 2-sticker arrays), as is illustrated in Fig. 4.

We fitted Model 2 to the data for the final 10 trials only. This model contained all experimental variables (but no interactions as they did not improve fit). There was a significant effect of age, feedback condition, and array size, as can be seen in Table 2. In the final 10 trials, children were more likely to produce a complex description if they were in the specific feedback condition or describing a target located in a 4-sticker array.²

Table 1
Fixed effects in Model 1

<table>
<thead>
<tr>
<th>Effect</th>
<th>B</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.374</td>
<td>0.348</td>
<td>-3.947</td>
<td>.0001</td>
</tr>
<tr>
<td>Age</td>
<td>-2.541</td>
<td>0.417</td>
<td>-6.092</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Time</td>
<td>0.048</td>
<td>0.009</td>
<td>5.343</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Feedback</td>
<td>0.519</td>
<td>0.406</td>
<td>1.279</td>
<td>.2010</td>
</tr>
<tr>
<td>Array</td>
<td>-0.130</td>
<td>0.235</td>
<td>-0.553</td>
<td>.5804</td>
</tr>
<tr>
<td>Age × Time</td>
<td>0.034</td>
<td>0.011</td>
<td>3.160</td>
<td>.0016</td>
</tr>
<tr>
<td>Time × Feedback</td>
<td>0.038</td>
<td>0.010</td>
<td>3.868</td>
<td>.0001</td>
</tr>
<tr>
<td>Time × Array</td>
<td>-0.016</td>
<td>0.010</td>
<td>-1.675</td>
<td>.0939</td>
</tr>
<tr>
<td>Feedback × Array</td>
<td>-0.345</td>
<td>0.195</td>
<td>-1.774</td>
<td>.0761</td>
</tr>
</tbody>
</table>

Note. Concordance between the predicted probabilities and the observed responses, \( C = 0.877 \). Somer’s \( D_{xy} \) (rank correlation between predicted probabilities and observed responses) = 0.754 (cf., Baayen, 2008).
2.3. Discussion

The results of this training study suggest that when children as young as 2½ years learn to become more informative, they do not simply become as descriptive as possible at all times. It also appears that receiving specific, contrastive feedback (containing models of appropriate referring expressions) in the right motivational context is more helpful than experiencing a more arduous process of repair (general feedback).

The specific feedback condition presumably rendered learning easier because the experimenter modeled helpful referring expressions (with relevant syntactic structures e.g., ‘‘I need the [person] [verb]ing’’) at a moment when the child was highly motivated to learn from them. Indeed this process of modeling expressions in settings where they are communicatively relevant has been shown to be effective for aphasic adults who possess the prerequisite social cognitive skills for communication but have lost the ability to produce the

Table 2
Fixed effects in Model 2

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$SE$</th>
<th>$z$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.063</td>
<td>0.412</td>
<td>0.153</td>
<td>0.878</td>
</tr>
<tr>
<td>Age</td>
<td>-1.519</td>
<td>0.478</td>
<td>-3.181</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Feedback</td>
<td>1.492</td>
<td>0.476</td>
<td>3.139</td>
<td>&lt;.002</td>
</tr>
<tr>
<td>Array</td>
<td>-0.827</td>
<td>0.179</td>
<td>-4.618</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note. Concordance between the predicted probabilities and the observed responses, $C = 0.910$. Somer’s $D_{xy} = 0.821$.
necessary expressions. In this population, too, language games similar to the ones used here have been shown to help both in encouraging language use and in inhibiting prepotent referring strategies (Davis & Wilcox, 1985; Pulvermuller et al., 2001).

Importantly, while training with specific feedback rapidly enabled children to describe stickers with complex descriptions, it did not lead them to use complex descriptions completely indiscriminately. Rather, children of both ages in both conditions tended to provide complex descriptions more often in contexts where they were required as opposed to contexts where a simple description would suffice. We make this conclusion cautiously because, as can be seen from Figs. 3 and 4, it is far from being the case that the children used complex descriptions only with 4-sticker arrays but rather they frequently underspecified on the 4-sticker array trials and overspecified on the 2-sticker array trials. We return to the issue of overspecification in the General discussion section.

The critical question now is what kind of knowledge this learning process has yielded and whether it would generalize. At the simplest level, we might ask whether children would have learned that ‘more distracters means you need to say more’ or whether children were assessing the similarity of distracters when choosing a referring expression. We address this question in Study 2.

3. Study 2

In this study, we trained and tested children by having them request stickers from 4-sticker arrays that either had similar distracters or dissimilar distracters (Fig. 5, top row). Children were always given specific feedback. We then tested for generalization to other array sizes by using 2- and 6-sticker arrays that either did or did not have similar distracters (Fig. 5, middle and bottom row). Of interest is whether children are still able to learn to be more informative and to adapt their referring expressions to an addressee’s need when contextual cues are more subtle and whether the array size affects this referring process. We are particularly interested to gather developmental data to address these questions as it has been shown that even adults only roughly gauge the complexity of a visual scene when deciding what referring expressions to produce (Arnold & Griffin, 2007; Fukumura, van Gompel, & Pickering, 2010; Kraljic & Brennan, 2005). We predicted that 2-year-olds would not have sufficiently mature selective attention to focus on the relevant parts of the more complex stimuli used in this study (Kloo & Perner, 2005) and therefore that they would be unlikely to learn to use complex descriptions in a discriminative fashion.

3.1. Method

3.1.1. Participants

Thirty-four normally developing, monolingual, English-speaking children were included in the study (19 boys, 14 girls), none of whom had participated in Study 1. There were sixteen 2-year-olds (mean age: 2;7, range: 2;5–2;9) and seventeen 4-year-olds (mean age: 4;0, range: 3;8–4;5). A further sixteen 2-year-olds participated but were excluded because they
did not complete the first testing session \((n = 11)\), did not attend the second test session \((n = 2)\), consistently asked for nontarget stickers \((n = 1)\), or turned out to be bilingual or not to have begun speaking \((n = 2)\). Testing took place at the Max Planck Child Study Centre, Manchester, UK. The children were predominantly White and middle class.

3.1.2. Materials and design

The picture book materials were exactly the same as for Study 1 except that the final sticker book, “The Bumbles go to the Farm,” had 12 pages (instead of 6) to allow sufficient items for a generalization test. Sticker boards were adapted as described below and illustrated in Fig. 5.

3.1.3. Procedure

The procedure was based on that of Study 1’s specific feedback condition, differing only in the type of sticker arrays used and the introduction of a generalization phase. The children thus took part in two test phases: training and generalization. During training, children were required to request 30 stickers from 4-sticker arrays with either similar or dissimilar distracters (see top of Fig. 5). Afterward, during generalization, children were required to request 12 stickers from either 2-sticker or 6-sticker arrays, again with either similar or dissimilar distracters (see middle and bottom of Fig. 5) and again given feedback when necessary.
Throughout the study, every other sticker was presented with similar distracters (counterbalanced such that each target sticker was seen an equal number of times with similar and dissimilar distracters for the experiment as a whole). In the generalization phase, both similarity and array size were counterbalanced and crossed such that, for the experiment as a whole, each target sticker was seen in each array size and with similar and dissimilar distracters an equal number of times. All children were given specific feedback when they did not provide sufficient information. The picture books were presented in a fixed order. Testing took place over two sessions (three new books made in each session) that were a maximum of 6 days apart (but, in the majority of cases, were on consecutive days).

3.1.4. Transcription and coding

The procedure for transcription and coding was identical to that for Study 1. Four subjects (two in each age group) were randomly selected and coded by a second coder, the first author. Coding agreement for first attempts was 100%.

3.2. Results

As with Study 1, children in both age groups rapidly began to verbally describe stickers after a few turns with feedback. The questions of interest are whether children would learn as easily as Study 1 and whether they would still be likely to produce complex descriptions more when they were necessary (i.e., when distracters were similar to the target) than when a simple description would suffice (distracters not similar) both during the training with 4-sticker arrays and later with 2- and 6-sticker arrays. To test this, we consider children’s first attempts at requesting a sticker during the 4-array phase (Fig. 6) and then consider the generalization phase (Fig. 7).

3.2.1. Training with 4-sticker array

As can be seen from Fig. 6, unlike Study 1, the 2-year-olds had difficulty in learning to produce complex referring expressions in this study. Further, although the 4-year-olds rapidly began to produce complex responses, they did not appear to adapt the complexity of their response to the similarity of the distractor stickers.

To test whether the children were at all sensitive to the need for more or less information, and whether this sensitivity varied as a function of age or time, we fitted a mixed-effects logistic regression model to the data for complex responses with child \((n = 34)\) as a random effect and age, similarity (of distracters), and time as fixed effects (adding two-way interactions did not improve fit significantly). The parameters of this model (Model 3) are reported in Table 3. As is to be expected, there were significant effects of age and time. However, there was no effect of similarity. That is, children were not more likely to provide more information when it was needed due to distractor similarity.

3.2.2. Comparison of training to Study 1

In addition to the lack of effect of similarity, another notable observation is that the children, particularly the 2-year-olds, appear to have been generally less likely to produce
Fig. 6. Frequency of complex responses as a function of age, time, and distracter similarity when the array size = 4. (Due to counterbalancing, the maximum possible frequency per time point is 8 for the 2-year-olds and up to 9 for the 4-year-olds.)

Fig. 7. Frequency of complex responses as a function of array size, age, and distracter similarity (maximum possible frequency per bar = 51).
complex responses than they were in the specific feedback condition in Study 1. We investigated this further by fitting a logistic regression model to data for complex responses collated from the two studies (including only data for 2-year-olds in the first 30 trials of the specific condition of Study 1 and during the 4-sticker training in Study 2). The model contained the variables time, study, and “contextual need for complex description” (i.e., whether or not a complex description would ideally be produced given the stimulus). The two-way interactions between study and time and between study and “need” were also included as they improved the fit. In addition to effects of study \((B = 1.714, SE = 0.741, z = 2.314, p = .021)\), time \((B = 0.130, SE = 0.168, z = 7.722, p < .0001)\), and need \((B = 0.885, SE = 0.257, z = 3.447, p < .0001)\), there was a study \(\times\) need interaction \((B = 0.749, SE = 0.362, z = 2.068, p = .039)\) and, critically, a significant study \(\times\) time interaction \((B = 0.081, SE = 0.023, z = 3.558, p = .0004)\). Thus, these younger children found it significantly harder to learn to produce complex descriptions at all in Study 2 and were less likely to produce them appropriately in response to their addressee’s needs. Similar results held for the 4-year-olds. This finding can be explained by the fact that, in Study 1, it was easier to see why more information was required to describe some targets than others. However, in study 2, it was less obvious why some trials required a complex description, whereas others resulted in success with a simple description. In this case, the 2-year-olds often failed to learn to provide complex expressions at all. Thus, at this early age, if children cannot identify the reason for switching between expression types in response to a given real-world contrast, then they will either fail to learn to use complex expressions at all or fail to use terms contrastively.

3.2.3. Generalization to other array sizes

In the generalization phase of Study 2, children continued to request stickers, but the number of distractors on the sticker board was changed to be either less than in the training phase (2-sticker array) or more (6-sticker array). Fig. 7 reports the number of times children produced a complex response (for their first attempt at reference) as a function of distracter similarity, array size, and age.

As can be seen, the 2-year-olds found it quite hard to produce complex expressions when they were required. The 4-year-olds were better able to do so and appear to have been more informative when the distracters were similar to the target. In addition, they appear to have reacted to array size by generally providing more information when the array was larger. To test the significance of these effects, we fitted a mixed-effects logistic regression model to

<table>
<thead>
<tr>
<th></th>
<th>(B)</th>
<th>(SE)</th>
<th>(z)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.669</td>
<td>0.365</td>
<td>-7.302</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Age</td>
<td>2.692</td>
<td>0.427</td>
<td>6.303</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Similarity</td>
<td>0.045</td>
<td>0.163</td>
<td>0.278</td>
<td>.781</td>
</tr>
<tr>
<td>Time</td>
<td>0.050</td>
<td>0.010</td>
<td>5.139</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Note. Concordance between the predicted probabilities and the observed responses, \(C = 0.87\). Somer’s \(D_{xy}\) (rank correlation between predicted probabilities and observed responses) = 0.765.

Table 3

Fixed effects in Model 3
the data on complex responses during the generalization phase, with age, array size, and distracter similarity and the two-way interactions between age and array size and between age and similarity as fixed effects (Note: Including a two-way interaction between array size and similarity and a three-way interaction did not significantly improve the fit). There were significant interactions of age \( \times \) array size \( (B = 1.242, SE = 0.515, z = 2.415, p = .0016) \) and age \( \times \) similarity \( (B = 1.118, SE = 0.516, z = 2.163, p = .031) \) but not other significant effects. We therefore fitted models to the data for 2-year-olds and 4-year-olds separately (in both cases, adding an interaction did not improve the fit). The model fitted to the 2-year-olds’ data revealed no effect of similarity nor array size \( (p's > .6) \), whereas the model fitted to the 4-year-olds’ data, Model 4, revealed effects of both factors as can be seen in Table 4.

### 3.2.4. Other response types

It is worth considering briefly what the children were doing during the generalization phase if they were not uniquely identifying referents. Fig. 8 shows the number of times children produced a simple or pointing response as a function of distracter similarity, array size, and age.

It is clear that whereas the 2-year-olds were just as likely to produce a simple response whether distracters were similar to the target or not, the 4-year-olds were less likely to do so when distracters were similar. This was confirmed when we fitted a mixed-effects logistic regression model to the data on simple responses during the generalization phase, with age, array size, and distracter similarity and the two-way interaction between age and similarity as fixed effects (Note: Model gave best fit). There was a significant age \( \times \) similarity interaction \( (B = 1.479, SE = 0.541, z = -2.732, p = .006) \).

When we consider pointing response’s it appears that the children’s tendency to point interacts with array size and distracter similarity such that when the array is small they are less likely to point if the distracters are similar to the target, but when the array size is large they are more likely to point if the distracters are similar to the target. This interaction between the number of distracters and their similarity was shown to be significant \( (B = 2.862, SE = 1.082, z = 2.645, p = .008) \) when we fitted a mixed-effects logistic regression model to the data on pointing responses during the generalization phase, with age, array size, distracter similarity, and the two-way interaction between array size and similarity as fixed effects (model with best fit). It would be interesting to further study whether children, 2-year-olds in particular, are actively choosing whether to point in response to a perceived need for information and inability to provide that information verbally.

<table>
<thead>
<tr>
<th>Fixed effects in Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B )</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>Array size</td>
</tr>
<tr>
<td>Similarity</td>
</tr>
</tbody>
</table>

Note. Concordance between the predicted probabilities and the observed responses, \( C = 0.898 \). Somer’s \( D_{xy} \) (rank correlation between predicted probabilities and observed responses) = 0.795.

Table 4
3.3. Discussion

Three main observations came out of Study 2. First, the comparison of learning in Studies 1 and 2 showed that children found it harder to learn to produce sufficiently complex descriptions in Study 2 than they did in Study 1. Second, despite the task being somewhat harder than in Study 1, the 4-year-olds nonetheless did produce complex descriptions and, by the time they had reached the generalization test, tended to provide complex descriptions in contexts where they were needed (i.e., when the target was of the same type as a
distracter) more than in those where they were not. Third, the sheer number of distracters affected the 4-year-olds responses such that they were more likely to produce a complex description if there were more items in the array regardless of the strict need for information. Array size also affected 2-year-olds’ pointing responses in interaction with distracter similarity (such that when the array was small they were less likely to point if the distracters were similar to the target, but when the array size was large they were more likely to point if the distracters were similar to the target). This is less straightforward to explain but underlines how the general properties of the visual scene influence children’s responses even if they are, strictly speaking, irrelevant. It could be, for example, that when there are a few but similar stickers, 2-year-olds “realize” that pointing would not work and try for a verbal description. In contrast, when they are overwhelmed with many similar distracters, they revert to pointing (see van der Sluis & Krahmer, 2007, for a discussion of the circumstances under which adults chose to point).

4. General discussion

The present studies were conducted to establish (a) what it is that children learn from receiving feedback in response to their attempts at reference and (b) what kinds of feedback are most helpful in promoting learning. We were particularly interested to know whether children learn to become more informative but do so at the cost of being as informative as possible at all times and whether this might occur more when children are given specific feedback. Study 1 found that specific feedback was more effective in increasing the rate at which children produced complex descriptions when they were required by an interlocutor. It also established that children do not appear to learn a general heuristic that they need to “be as descriptive as possible at all times.” Even in the final trials of Study 1, once children had learned to produce a large number of complex descriptions, they tended to produce more information when it was required than when it was not. In this study, it was relatively easy for children to establish how much information they needed to give their addressee on the basis of array size. In Study 2, we raised the stakes by holding array size constant during training and varying how similar the distracters were to the target. In this case, the 2-year-olds found it far harder to learn. They provided fewer complex descriptions than in Study 1 and, when they did so, it was without regard for distracter similarity. Apparently, the “distracter similarity” cue to potential ambiguity was too opaque for such young children. In contrast, the 4-year-olds did produce complex descriptions and tended to do so more when it was required due to the distracters’ similarity to the target. Interestingly, during the generalization phase, the 4-year-olds also tended to give more information when the array size was large than small, an observation that we return to below.

The above results support an account whereby children learn to produce complex referring expressions by collaboratively building up good descriptions and remembering them such that, in the future, when contexts are deemed to be sufficiently similar, reference can proceed more smoothly. Critical to this process is the ability to conceptualize the target and context in terms of features that are relevant to the communicative endeavor. When it is easy
for children to do this and discern why more information is required to describe some targets than others, even young children rapidly learn to provide extra information when it’s needed and not when it would be redundant (as was observed in Study 1). However, if children cannot identify the reason for switching between expression types in response to a given real-world contrast, then they will either fail to learn at all or fail to use terms contrastively. Children’s ability to detect relevant contrasts in the conversational setting and to connect them with speakers’ intentions and listeners’ needs will become refined with development such that, by adulthood, people are quite proficient in rapidly learning to tailor referring expressions to context (Guhe & Bard, 2008).

Of course, the learning we observed here occurred in an artificial setting where the contexts from one trial to the next had clear parallels and thus it was relatively easy to retain successful strategies in the form of semi-abstract scripts (in the sense of Dale & Reiter, 1995; note that each target was unique, so scripts could not be learned verbatim but would need to be abstract). We assume that, in everyday life, some conversational settings will be similarly repetitive and can thus run on relatively fixed scripts, whereas others will be more demanding (e.g., when communicating with new people, about new things in new locations) and will rely more on laborious planning and negotiation (or else parental scaffolding) to be successful. It is worth noting that the fact that the use of a referring expression in a given context is practiced and scripted need not mean that it is devoid of audience design. These two qualities are somewhat orthogonal (Goudbeek & Krahmer, 2010). That is, it should be possible to have fairly fixed scripts that take account of some basic addressee information (perhaps simply whether the interlocutor is already attending to the object in question; Brennan & Hanna, 2009) and others that do not. Thus, the language system likely has a number of means of arriving at a good referring expression that are more or less automatic and more or less audience designed. What is critical is that this system is not static. As people accrue experience with similar situations, their piecemeal referring strategies will be merged and ‘redescribed’ (Karmiloff-Smith, 1992) to yield ever more effective means of guiding an addressee’s attention (although see Horton & Spieler, 2007, for the effects of aging on memory systems that support audience design).

This opportunistic learning process gives rise to considerable individual differences in referring style (Viethen & Dale, 2009). Nonetheless, we should expect to be able to find a number of stable properties in the environment that allow most people to derive similar referring heuristics. In the current study, one very clear example was 4-year-olds’ tendency to produce more complex expressions if there were more distracters in the visual scene (regardless of their similarity to the target). Although it often leads to overspecification, this is a reasonable heuristic to follow both from the point of view of the speaker and the addressee (whose time in visual search will decrease; for relevant adult data, see Arts, Maes, Noordman, & Jansen, 2011; Paraboni, van Deemter, & Masthoff, 2007).

There is already considerable evidence that adults also rely on partially valid, global cues to accessibility and do not necessarily attempt to produce referring expressions that convey all and only the information minimally required to uniquely identify a referent (e.g., Arnold & Griffin, 2007; Bard et al., 2007; Englehardt, Bailey, & Ferreira, 2006; Fukumura, van Gompel, & Pickering, in press; Heller, Skovbroten, & Tanenhaus, 2009; Kraljic & Brennan,
This is reflected to a degree in many computational linguistic algorithms (e.g., Dale & Reiter, 1995; Krahmer & Theune, 2002; Krahmer & van der Sluis, 2003; Paraboni et al., 2007; Viethen & Dale, 2008; Walker, 1996). From a developmental point of view, the fact that we use heuristics can help explain how the language system gets up and running while there is still significant linguistic and social cognitive development to be done (e.g., Moll, Carpenter, & Tomasello, 2007). The point here is that children’s developing language skills will to some extent be shaped by their inability to rapidly and explicitly reason about potential ambiguities based on subtle cues to accessibility for an addressee (see O’Neill, 1996 for discussion). Fortunately, many heuristics need not rely on such an ability, particularly not at the point of use. Let us take the example of the 4-year-olds’ heuristic to say more when there are more stickers. It is easy to see how this could be used without any explicit audience design. However, in order to have learned this heuristic, one needed to engage with an interlocutor, to attempt to be understood, and to repair. In this sense, audience design of some form is involved in the learning of the heuristic if not in its use. This learning process is not without consequence for psycholinguistic and computational linguistic models: It would not be surprising if the communicative strategies children learn early on remained in some form even in the adult processing system. That is, any model of human communication should be seen as the product of a protracted, indeed continuing, developmental process.

Limited computational linguistic attempts to reflect this learning process have been made by augmenting referring expression generation (REG) algorithms with statistical information and employing machine learning (Jordan & Walker, 2005; Viethen, Dale, Krahmer, Theune, & Touset, 2008). It is not currently clear whether these approaches would benefit from further emulating the human learning process we have described here. One serious limit on the feasibility of doing so is due to the large number of interdependent problems being simultaneously solved by the child. For example, perhaps the most difficult challenge that faces the child learner is to perceive the visual scene in terms of relevant features/attributes in the first place. Importantly, a number of studies have shown that comprehension of referring expressions shapes children’s perception of potential referents and promotes the formation of categories (e.g., Balaban & Waxman, 1997; Namy & Gentner, 2002; Plunkett, Hu, & Cohen, 2008). Thus, children’s experience of referential communication will teach them both about attributes and about referring expressions.

As well as drawing children’s attention to the structure of the environment and the function of referring expressions, it is likely that the current training study also made the process of communication itself more transparent for young children. When giving specific feedback, the experimenter made it very clear why she needed more information by highlighting that both the target and a distracter met the child’s description. This feedback was particularly effective in Study 1, where the arrays that called for more complex descriptions were markedly different from those that did not. In this case, multiple cues to accessibility converged. We would argue that encountering such situations where it is relatively straightforward to detect the function of referring expressions is critical not just to learning the function of referring expressions but also to developing metalinguistic insight. Indeed, there is good reason to believe that it is by engaging in conversations (especially, those
highlighting differences in perspective) that children learn about other minds (Astington & Baird, 2005; Ensor & Hughes, 2008; Lohmann & Tomasello, 2003; Nelson et al., 2003; Peterson & Siegal, 1999).

To summarize, we propose that even the youngest children in the current study will have brought a range of nascent social–cognitive skills to the communicative situation. However, these skills are unlikely to be fully adult-like and, thus, the informational needs of the adult interlocutor are not likely to have been fully transparent, especially to the 2-year-old participants. Nonetheless, with experience of feedback and repair, even 2-year-olds became able to request novel targets in a relatively sophisticated manner when cues to accessibility were strong (in Study 1). The advantage of receiving more specific feedback demonstrates that, like adults, children are able to make the most of occasions where an addressee indicates what type of expression is needed for an attempt at reference to be successful (Clark & Wilkes-Gibbs, 1986). We propose that it is by engaging in such processes of repair that children build up a repertoire of conversational models, learn about relevant features of their environment, gain insight into other minds, and ultimately come to fully understand why the information they provide in conversation is needed in the first place.

Notes

1. We might also ask how speakers should meet these requirements if they want to secure and speed comprehension on the part of their addressee. This is quite a different question because what speakers actually say in a task is not always what is easiest for a listener to understand, and what listeners judge to be an optimal description is not necessarily the description that is easiest for a listener to act on (Belz & Gatt, 2008).

2. Although Model 2 did not contain interactions, it is of particular interest to us whether the 2-year-olds were able to provide more complex descriptions when they were required (4-sticker array) than when they would have been overspecifications (2-sticker array). We therefore fitted a model to the data for the 2-year-olds’ final 10 trials separately, including array and feedback condition as fixed effects (but not the interaction between the two as it did not improve fit). For this model, there was also an effect of array ($B = -0.6759, SE = 0.260, z = -2.604, p = .009$).

Acknowledgments

The authors would like to thank Lianne Heys and Elizabeth Wills for help with data collection; Siu-lin Rawlinson for drawing the story books; Colin Bannard, Tanya Behne, Stephen Butterfill, Malinda Carpenter, Dan Carroll, Jane Herbert, Grzegorz Krajewski, and Aylin Küntay for helpful discussions; Jette Viethen for directions through the REG literature; and our volunteer families, Freshfield Nursery School, and Kids Unlimited Didsbury for participating. This research was funded by the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
References


