Developmental Psychology
2003, Vol. 39, No. 4, 717–729

Copyright 2003 by the American Psychological Association, Inc.
0012-1649/03/$12.00 DOI: 10.1037/0012-1649.39.4.717

The Influence of Language and Socioeconomic Status on Children's Understanding of False Belief

Marilyn Shatz
University of Michigan

Gil Diesendruck
Bar-Ilan University

Ivelisse Martinez-Beck
U.S. Department of Health and Human Services

Didar Akar
Boğaziçi University

Study 1 investigated whether differences in the lexical explicitness with which languages express false belief influence children's performance on standard false belief tasks. Preschoolers speaking languages with explicit terms (Turkish and Puerto Rican Spanish) were compared with preschoolers speaking languages without explicit terms (Brazilian Portuguese and English) on questions assessing false belief understanding either specifically (the *think* question) or more generally (the *look for* question). Lexical explicitness influenced responses to the *think* question only. Study 2 replicated Study 1 with groups of both speakers differing in socioeconomic status (SES). A local effect of explicitness was found again as well as a more general influence of SES. The findings are discussed with regard to possible relations among language, SES, and understanding of mind.

Research on children's understanding of mind has resulted in a profusion of reports about the nature of and influences on the development of false belief understanding, a capacity considered central to mature social thought and behavior (e.g., Lewis & Mitchell, 1994; Perner, 1991; Wellman, 1990). Although one can question whether understanding of false belief should be taken as paradigmatic for an understanding of mind (Bloom & German, 2000), nonetheless a focus on false belief understanding has dominated the field since Wimmer and Perner (1983) initiated research into the area by using an experimental task to show that preschoolers had difficulty predicting the actions of others on the basis of false belief states. The original findings are remarkably robust (see Wellman, Cross, & Watson, 2001, for a review). Despite studies reporting various factors that facilitate good performance on false belief tasks, children under the age of 4 years generally have trouble on tasks requiring them to predict the thoughts or actions of others that are based on false beliefs. And a close look at the findings reveals that there is by no means a ceiling effect on performance even at age 4. Such robustness is compatible with suggestions that there is a universal cognitive developmental trajectory governing growth in false belief understanding. Obviously, cross-cultural research showing common age and stage patterns of development would support a universal cognitive theory, and there is some such support (see Avis & Harris, 1991, and Lillard, 1998, for discussions).

One cognitive capacity that has been tied to false belief understanding is language. Various language abilities such as vocabulary size and use of certain syntactic structures have been shown to be related to understanding of mind (see Astington & Jenkins, 1999, and deVilliers, 2000, for reviews), as have family talk about mental states (e.g., Dunn, Brown, & Beardsall, 1991) and bilingualism (Goetz, 2003). Despite the tantalizing data indicating a relation between language and false belief understanding, the nature of the relation remains unclear. Thus, further exploration of the role of language is warranted, especially with regard to any influence cross-linguistic differences may have in the development of such understanding (see Vinden, 1996, for a discussion).

More particularly, we asked whether differences in the lexical explicitness with which languages express false belief might influence children's ability to pass false belief tasks. Examining the possible influence of lexical explicitness within only one language, Lee, Olson, and Torrance (1999) found that Chinese-speaking preschoolers did better on a false belief task when the mental verb used in the task explicitly connoted false belief than when the verb had a neutral connotation as to the truth of the belief. We instead explored the issue cross-linguistically with speakers of four different languages. Would the glimmers of understanding that English-speaking 3-year-olds show in modified tasks be more developed in same-aged children whose native languages have

Marilyn Shatz, Department of Psychology, University of Michigan; Gil Diesendruck, Department of Psychology, Bar-Ilan University, Ramat Gan, Israel; Ivelisse Martinez-Beck, Child Care Bureau, Administration for Children and Families, U.S. Department of Health and Human Services, Washington, DC; Didar Akar, Department of Linguistics, Boğaziçi University, Istanbul, Turkey.

The order of the second and third authors is merely alphabetical.

This research was supported by a grant from the Office of the Vice President for Research, University of Michigan. Ivelisse Martinez-Beck held a National Science Foundation predoctoral fellowship. Portions of Study 1 were presented at a symposium on Language and Thought at the biannual meeting of the Society for Research in Child Development, April 1995, Indianapolis, Indiana, and portions of Study 2 were presented at Head Start's Fourth National Research Conference on Children and Families, July 1998, Washington, DC.

Correspondence concerning this article should be addressed to Marilyn Shatz, Department of Psychology, 525 East University Avenue, University of Michigan, Ann Arbor, Michigan 48109-1109. E-mail: mshatz@umich.edu

linguistic means of encoding false belief more explicitly—and would they be developed enough that such children could pass standard false belief tasks earlier than children without such terms in their language? Or, despite language differences in the lexical explicitness of expressing false belief, would the difficulties young children experience with false belief tasks persist across groups speaking different native languages? To address these questions, we examined the performances on false belief tasks of 3- and 4-year-old monolingual native speakers of four languages. Two of the languages, Turkish and Puerto Rican (PR) Spanish, have ways to mark false belief states explicitly with specific verbal forms, whereas the other two languages, English and Brazilian (BR) Portuguese, have no such specific forms. If native speakers of Turkish and PR Spanish were to do better on the false belief tasks, that would be evidence against the idea that there is a universal cognitive limitation on the early understanding of false beliefs and in favor of the idea that the lexical characteristics of languages can influence performance on this cognitive task, at least with regard to the age at which such a cognitive limitation may be overcome.

In what follows, we briefly describe the relevant differences among the four languages, ask the questions that we investigated with speakers of these languages, and note which groups of speakers were used to address each of them.

The Varied Expression of Belief

Languages need to express a variety of meanings about belief with mental terms, but often the same term is used to convey more than one meaning. Consider the English word *think*, which is commonly used to report the performance of a mental action ("I'm thinking about the weather"); to report a belief when the speaker is neutral about the truth value of the belief ("John thinks it will rain today"); as well as to report a belief when the speaker knows the belief is false ("John thinks Richard Nixon is alive"). Turkish and PR Spanish both have specific forms for marking false belief. Turkish uses the verbs *san* (a Turkish root) and *zannet* (from the Arabic), with *san* reportedly being the preferred form among the younger generations. *Düşün* is the verb form used when the speaker is neutral on truth value, and this form can also mean performing the mental activity of thinking. (Turkish also has various alternative forms to indicate belief states that are not discussed here.) PR Spanish uses *creer* for the neutral truth-value cases; when the speaker is sure that a false belief is held, the reflexive clitic is added to the verb phrase, resulting in the infinitive form, *creer-se*. The clitic can be positioned separately from the verb in the sentence, but it still depends on the occurrence of the verb. There is a different term, *pensar*, for the meaning of performing mental action. Like PR Spanish, but unlike English, BR Portuguese uses the verb *pensar* for performing mental action. Unlike PR Spanish but like English, BR Portuguese uses only one verb construction for believing in either the neutral or false case, namely *achar*. However, unlike the English verb *think*, in certain syntactic constructions, *achar* can instead mean to find. (See the Appendix for a fuller description of how belief is encoded in each language.) Table 1 summarizes the terms commonly used to express, in the four languages, the three meanings of engaging in the act of thinking, thinking neutrally with regard to truth value, and believing falsely.

Does Explicit Linguistic Marking Influence False Belief Understanding?

The differences among these four languages allow us to investigate four questions about the relation between language and the understanding of mind. First, does having an explicit term for false believing in one's language improve performance on false belief tasks? If so, other things being equal, Turkish and PR Spanish speakers should do better than BR Portuguese and English speakers.

Second, how might lexically explicit forms for false belief help? On the one hand, if they help by focusing the children on the representation of belief and encouraging them generally to reason about representation's relation to reality, then Turkish and PR Spanish speakers in false belief contexts should do as well when asked a general question about false belief (e.g., Where will x *look for* the y?) as when asked an explicitly marked question (e.g., Where does x *think* the object is?). Moreover, they should do as well on such questions when the speaker uses a neutral or unmarked form as when the marked form is used, because the mere existence of the marked form in the language would have previously facilitated their ability to consider a false representation–reality relation irrespective of the language used in any particular case. On the other hand, if the marked form helps only locally without influencing reasoning in a more general way, then the Turkish and PR Spanish speakers should perform better when the marked form is actually used than when the neutral, unmarked form is used or when a question lacking the explicit form is asked.

Third, do the two kinds of marking for false belief—main verb in Turkish and addition of the clitic in PR Spanish—influence performance equally? If so, there should be no differences between Turkish and PR Spanish speakers. However, if one kind of marking is more discriminable or salient than the other, then we would expect differences between the two groups of speakers.

Table 1
Terms for Three Meanings of Think in Four Languages

Meaning	Turkish	Puerto Rican Spanish	Brazilian Portuguese	English
Perform mental action	düşün*	pensar	pensar	think
Believe (neutral)	düşün	creer	achar	think
Believe (falsely)	san* zannet*	creer-se	achar	think

Note. Asterisks indicate root forms.

Fourth, does the differential polysemy of *think* terms in BR Portuguese and English influence performance? In particular, perhaps the BR Portuguese speakers hearing *achar*, which under some circumstances can carry the factive meaning of to find, may be negatively influenced by that term's potential to carry such a meaning, especially if they are asked a general *look for* question before the *achar* question. Hence, their responses might be sensitive to question order, whereas the responses of English speakers hearing *think*, without such a potential for polysemy, should show no sensitivity to question order.

In Study 1, we present data relevant to all four of these questions, with each data set making use of the same set of tasks but different combinations of participants. First we used speakers of all four languages to address Questions 1 and 2, and then we used speakers of Turkish and PR Spanish to provide more evidence regarding Question 2 and also to address Question 3. Then we used English and BR Portuguese speakers to address Question 4 about the latter group's sensitivity to the polysemy of *achar* and the order of false belief questions. In Study 2, we used new groups of English and PR Spanish speaking children to follow up on a possible influence of socioeconomic class differences in false belief understanding that was suggested by Study 1 data.

Study 1

Method

Participants

Two hundred thirty children participated in this study: 113 three-year-olds (mean age = 3 years 6 months; range = 3 years 0 months to 3 years 11 months) and 117 four-year-olds (mean age = 4 years 5 months; range = 4 years 1 month to 4 years 11 months). There were 50 Turkish children (24 girls and 26 boys from Istanbul), 60 Puerto Rican children (34 girls and 26 boys from San Juan), 60 Brazilian children (28 girls and 32 boys from Sao Paulo), and 60 American children (28 girls and 32 boys from two small midwestern cities). Twenty-three of the Turkish children were 3-year-olds; the remaining three language groups were split evenly between the two age groups. Preliminary analyses on each of the two age groups revealed no differences in age among the language groups.

All of the children were native monolingual speakers of their respective languages and attended preschools in their respective countries. To avoid possible differences that were due not to language but to socioeconomic differences, we limited our choices of preschools either to university-related or private ones in the hope that relatively homogeneous samples could be drawn from them. The Turkish children were drawn from four private preschools and were mostly of an upper-middle-class socioeconomic status (SES); the Puerto Rican children attended a state-university-affiliated preschool and ultimately were revealed to be more diverse in SES; the Brazilian children attended a private preschool and were of upper-middle-class SES; and the American children were drawn from a state-university-affiliated preschool and a private preschool and were mostly of upper-middle-class SES.

Design

The study was designed to allow for three comparisons, with each one using a subset of the participants (see Table 2 for a summary). The first comparison involved 118 children from the four language groups in a 2 (age: 3-year-olds or 4-year-olds) × 4 (languages: Turkish, PR Spanish, BR Portuguese, or English) between-subjects design. The purpose of this comparison was to test for the effect of explicit language marking of false belief on children's performance in false belief tasks by comparing languages with marking (Turkish and PR Spanish) and without marking (BR Portuguese and English). For this comparison we planned to use 27 Turkish speakers (12 three-year-olds and 15 four-year-olds) and 31 PR Spanish speakers (16 three-year-olds and 15 four-year-olds), who heard in the task protocols the verbal forms marked for false belief in their respective languages, and 60 BR Portuguese and English speakers (15 in each age/language group), who heard the same false belief questions in the same order as the Turkish and PR Spanish speakers.

The second comparison involved all 110 of the Turkish- and PR Spanish-speaking children in a 2 (age: 3-year-olds or 4-year-olds) × 2 (condition: marked or unmarked) × 2 (languages: Turkish or PR Spanish) between-subjects design. The purpose of this comparison was to use within-language comparisons to explore the generality of any effect of explicit marking of false belief on the children's performance, as well as to examine the relative value of main verb and clitic markings. The conditions differed only in the versions of verbal forms used: The standard forms for false belief in Turkish and PR Spanish (*san* and *creerse*, respectively) were used in the marked condition; the neutral forms (*düşün* and *creer*) were used in the unmarked condition. Twelve 3-year-old and 15 four-year-old

Table 2
Design of Study 1

Comparison	Language of participants and condition	Issues addressed
1	Turkish and PR Spanish—marked BR Portuguese and English—unmarked	Does marking influence false belief performance across languages? Is the influence general? Does marking interact with age?
2	Turkish and PR Spanish—marked vs. unmarked	Does marking influence false belief performance within languages? Are main verb and clitic markings equivalent in influence?
3	BR Portuguese and English—*think* question first vs. *think* question second	Does polysemy of the *think* term influence false belief performance? Does order differentially affect BR Portuguese and English?

Note. All comparisons included groups of both 3- and 4-year-old participants, and all included both *think* and *look for* questions.

Turkish speakers and 16 three-year-old and 15 four-year-old PR Spanish speakers were randomly assigned to the marked condition; 11 three-year-old and 12 four-year-old Turkish speakers and 14 three-year-old and 15 four-year-old PR Spanish speakers were randomly assigned to the unmarked condition.

The third comparison involved 120 BR Portuguese- and English-speaking children in a 2 (age: 3-year-olds or 4-year-olds) × 2 (order: *think* question first or *think* question second) × 2 (languages: BR Portuguese or English) between-subjects design. The purpose of this comparison was to investigate a possible influence of a polysemy differential in the *think* terms as well as to examine the relation between polysemy and the order of questions asked. Here we compared the performance of the two nonexplicit language groups, using the identical tasks but reversing the orders of the *think* and *look for* questions. Half of the children in each language/age group were given each order (15 BR Portuguese speakers and 15 English speakers in each age/order group).

Tasks and Materials

There were two moved-objects tasks and two surprise-contents tasks similar to those used in classic false belief research. The moved-objects tasks were presented via storybooks with line drawings of the characters and objects. Each story involved the placement, unbeknownst to the main character, of a familiar object: a juice box that was removed from the refrigerator and placed elsewhere and a ball that had been moved from under a chair and rolled under a bed. One surprise-contents task involved a clearly identifiable crayon box and an equally sized and shaped box covered in solid blue paper. The other involved a clearly identifiable Lego box and an equally sized and shaped box covered in solid pink paper. Crayon and Lego boxes had been selected because the researchers native to the participants' countries judged them to be familiar to children from these countries.

The order of presentation of the tasks was the same for all children: First the crayon box surprise-contents task, then the juice and the ball moved-objects tasks, and last the Lego box surprise-contents task. A single order was determined on the basis of pilot data, which revealed that it was easier to engage the children by starting with a moved-objects task and ending the session with the Lego box task because the children wanted to continue to play with the Legos.

Procedure

All children were given all four tasks. Children were seen individually by two experimenters in a single session. Experimenter 2 (E2) brought the child into a room where Experimenter 1 (E1) was waiting, sitting by a small table. The child and E2 were invited to sit by the table "to play some games." E1 showed the child a bag "that has lots of things to play with." E1 then pulled out of the bag the crayon box and the blue box. Immediately, E2 touched the crayon box and said that he or she wanted to draw with the crayons and was going to get some paper. E2 then left the room. E1 then asked the child, "What box do you think has the crayons in it?" Following the child's answer, E1 opened both boxes and remarked on the fact that the crayon box was empty but the blue box contained crayons. The child was then asked the following two questions:

1. "Where does [E2] think the crayons are?"

2. "Where is [E2] going to look for the crayons when [he/she] returns to draw?"[1]

E2 then returned to the room and approached the table, pulled the crayon box toward her or him, and stated that she or he could not find the drawing papers in the other room. E2 went to the corner of the room, facing the wall, to look for the papers in her or his bag. The child was asked the following two questions:

3. "Why did [E2] pick/grab the crayon box?" (We asked this question to elicit explanations of E2's behavior.)

4. "Where are the crayons really?" (This question was asked as a reality check. The child was to tell E2 where the crayons were when E2 returned to the table with the drawing papers. The child and E2 were then left to draw for a few minutes.)

After the child had drawn with the crayons, E1 pulled out of her bag a storybook with the moved-objects stories. After each story, the child was asked Questions 1–4 above with the appropriate changes; that is, the questions were about the main characters in the stories instead of E2, and they were about moved objects and not about which box held crayons.

When E1 finished getting answers to the story-related questions, she pulled out of her bag the Lego box and the pink box. At this point, E2 returned to the table, touched the Lego box, pulled it towards her or him, and told E1 that she or he was going to play with the Legos but first had to go clean up the papers and crayons and also look for some other Legos. E2 then left the room, and the procedure continued in a manner similar to that described for the crayon task. At the end of the session, the child and the two experimenters played with the Legos. The sessions took about 20 min.

Each participant heard only either a marked or unmarked version of the *think* questions in the set of questions described above, but each child was asked all four questions for each of the tasks. For the unmarked version, the neutral verb not explicitly expressing false belief in Turkish and PR Spanish and the standard verb in BR Portuguese and English were used in the *think* questions; for the marked version, the marked verb that expresses false belief in Turkish and PR Spanish was used in the *think* questions. Each participant heard only one order of questions. The *think* first version, in which the *think* question preceded the *look for* question for each task, was used throughout the first two comparisons and for half the children in the third comparison, and the *think* second version, in which the *look for* question preceded the *think* question for each task, was used for half the children in the third comparison.

Coding

Questions 1, 2, and 4. A correct response to each trial of the *think*, the *look for*, and the reality questions in the four tasks was coded as a 1; all other responses were coded as 0. Composite scores from 0 to 4 for each of Questions 1, 2, and 4 were then assigned to each child by adding his or her responses across the four tasks.

Question 3. Children's explanations of the characters' actions were coded in seven different categories as involving (a) characters' beliefs, (b) lack of knowledge, (c) information seeking, (d) actions or perceptual characteristics, (e) purpose, need, or desire, (f) story repetitions or reality statements, or (g) irrelevant statements. Intercoder reliability for children's justifications was 89%. See Table 3 for examples of the explanations used by the children in each category.

Results

We begin with a preliminary analysis of the children's responses to Question 4, which served as the reality check. We then present analyses of the Comparison 1 data, comparing the responses of

[1] Another question—"Is [E2] going to find the crayons here [where the child had pointed in reply to the previous questions]?"—was asked at this point. It was included to assess the consistency of the children's answers with their answers to the prior questions, and the answers were coded accordingly, but the data are not reported here. We therefore numbered the questions analyzed and reported here as 1–4 for ease of exposition.

Table 3
Examples of Explanations Given in Response to Question 3 in Study 1

Explanation	Language	Age
Characters' beliefs		
"Ella se cree que esta ahi." [She believes (falsely) that it is there.]	PR Spanish	4 years 5 months
"Because he thinks there's crayons in there."	English	3 years 11 months
Lack of knowledge		
"He doesn't know the crayons are not in there."	English	4 years 3 months
"No sabe que no esta ahi." [She doesn't know that it is not there.]	PR Spanish	4 years 1 month
Information seeking		
"He has to find it."	English	3 years 8 months
"Esta buscando y no lo va a encontrar." [She is looking and she is not going to find it.]	PR Spanish	4 years 7 months
Action/perceptual		
"It has pictures of crayons in it."	English	4 years 9 months
"Porque la puso ahi." [Because he put it there.]	PR Spanish	4 years 2 months
Purpose, need, or desire		
"He wants to throw it."	English	3 years 2 months
"Porque tiene sed." [Because she is thirsty.]	PR Spanish	4 years 7 months
Reality statements		
"Her brother finished it."	English	4 years 3 months
"Porque ahi no esta." [Because it is not there.]	PR Spanish	3 years 7 months
Irrelevant		
[The child only laughs]	English	3 years 8 months
"I don't know."	English	4 years 3 months

Note. PR Spanish = Puerto Rican Spanish.

children who speak a marked language (Turkish and PR Spanish) and who heard such markings with those of children who speak an unmarked language (BR Portuguese and English) on first the *think* question and then the *look for* question. We next present for both of these questions the analyses of the Comparison 2 data, comparing the responses of the Turkish and PR Spanish speakers who heard the marked version of their languages with those of the Turkish and PR Spanish speakers who heard the unmarked versions of their languages. We also address here whether one sort of marking is preferred over another. For Comparison 3, we compare the performance of BR Portuguese and English speakers who heard the *think* question first with the performance of those who heard it second. Finally, we present data on the children's explanations of their choices.

Preliminary Analysis—Question 4

We began our analyses by examining the children's responses to the last questions in each task, which served as reality checks. If children could not answer these questions adequately, we could not be sure how to interpret their answers on the other questions with regard to their false belief understanding (see Wellman et al., 2001, for more on this point). Therefore, we resolved to eliminate from further analyses any children who had not achieved a score of 3

or 4 on Question 4. We found that only one half of all the PR Spanish speakers (9 three-year-olds and 21 four-year-olds) achieved such a score, whereas 90% of all Turkish speakers (19 three-year-olds and 26 four-year-olds), 92% of all BR Portuguese speakers (26 three-year-olds and 29 four-year-olds), and 75% of all English speakers (18 three-year-olds and 27 four-year-olds) did so. The relative frequencies of speakers eliminated (30 PR Spanish, 5 Turkish, 5 BR Portuguese, and 15 English speakers) were significantly different across languages, $\chi^2(3, N = 230) = 35.80$, $p < .01$.

In discussing why the PR Spanish speakers might have done more poorly than the other speakers, their experimenter noted that the strategic criterion of selecting a university-related preschool or a school in a middle- to upper-middle-class neighborhood to assure participant homogeneity in SES may have been less successful in Puerto Rico, because there had seemed to her to be considerable heterogeneity in the university-related classroom. Following up on this observation, we solicited from the school's records information on the children's socioeconomic backgrounds. However, the only information available to us post hoc was the parental occupations of the PR Spanish speakers. Therefore, we adopted the following strategy. For the present study, we coded primary parents' occupations according to the 1989 Socioeconomic Index of Occupational Prestige (SEI; Nakao & Treas, 1992) and eliminated from further analysis all PR Spanish-speaking participants whose parental occupations were rated below 50 on the SEI. This left 17 children (5 three-year-olds and 12 four-year-olds) whose parents worked in jobs requiring more than 12 years of education or training in a trade, thereby likely making them more similar to the children in the other groups. Table 4 shows the numbers of participants by language, age, and condition who remained after elimination because of Question 4 performance or SES and whose data were used in the statistical analyses in this section. We also conducted a subsequent study to address directly the question of SES (see Study 2).

Table 4
Numbers of Children Included in the Final Analyses by Age, Language, and Condition in Study 1

	Condition		
Language and age	Marked	Unmarked	*Think* question second
Turkish			
3-year-olds	8	11	—
4-year-olds	14	12	—
PR Spanish			
3-year-olds	2	3	—
4-year-olds	7	5	—
BR Portuguese			
3-year-olds	—	14	12
4-year-olds	—	15	14
English			
3-year-olds	—	9	9
4-year-olds	—	14	13

Note. Table shows the numbers of children remaining after dropping children (a) who responded to fewer than three reality questions correctly and (b) whose parents' Socioeconomic Index of Occupational Prestige scores were less than 50. Dashes indicate that no children were tested in those conditions. PR = Puerto Rican; BR = Brazilian.

Comparison 1: Marked–Unmarked Four-Language Analyses

We addressed whether there was any influence of explicitly marking false belief in a language on children's performance on false belief tasks by analyzing first whether an effect was found on responses to questions directly asking about thinking and then whether a more general effect was found on responses to the *look for* question. Table 5 presents the means on the *think* and *look for* questions for the subjects remaining in each of the language/age groups. Because so few PR Spanish speakers remained, we combined the data from Turkish and PR Spanish speakers into a marked language condition and the data from the BR Portuguese and English speakers into an unmarked language condition for the statistical analyses.

Question 1—the think *question.* A 2 (condition: marked language or unmarked language) × 2 (age: 3-year-olds or 4-year-olds) repeated measures analysis of variance (ANOVA) for unequal samples revealed a significant effect of age, $F(1, 79) = 35.73$, $p < .01$, with 4-year-olds answering more *think* questions correctly ($M = 2.94$) than 3-year-olds ($M = 1.42$); a marginal effect of condition, $F(1, 79) = 3.93$, $p = .05$; and no interaction effect. In fact, analyses against chance (chance = 2) revealed that 3-year-olds in the unmarked condition ($M = 1.35$) correctly answered significantly fewer questions than expected by chance, $t(22) = -2.92$, $p < .01$; 3-year-olds in the marked condition ($M = 1.6$) were correct at chance levels, $t(9) = -1.08$, $p > .3$; and 4-year-olds in both the unmarked ($M = 2.62$) and marked ($M = 3.38$) conditions correctly answered significantly above chance, $t(28) = 2.99$, $p < .01$ and $t(20) = 6.50$, $p < .01$, respectively.

With no interaction between condition and age, we combined across ages and performed a one-way, repeated measures ANOVA for unequal samples. It revealed a significant effect of markedness, $F(1, 81) = 6.60$, $p < .05$. Speakers of the languages explicitly marking false belief, Turkish and PR Spanish, answered the *think* question correctly more often than did the speakers of unmarked languages, BR Portuguese and English (*M*s = 2.81 and 2.06, respectively). Analyses against chance revealed that children in the marked condition performed significantly above chance, $t(30) = 3.38$, $p < .01$, whereas children in the unmarked condition performed at chance level, $t(51) = 0.33$, $p > .7$.

We also conducted a chi-square test on the number of children (combined across ages) who got 3–4 *think* questions correct. Seventy-one percent of the children hearing marked forms in their languages did so, compared with only 37% of the children who speak unmarked languages, $\chi^2(1, N = 83) = 9.21$, $p < .01$.

Question 2—the look for *question.* The same 2 × 2 repeated measures ANOVA conducted above but now with the *look for* question as the dependent variable yielded only a main effect of age, with older children performing more correctly ($M = 2.88$) than younger children ($M = 1.67$), $F(1, 79) = 19.44$, $p < .01$. The repeated measures ANOVA combining across ages on this question revealed no effect of markedness ($p > .3$). Although 65% of children in the marked condition got 3–4 *look for* questions correct, compared with 46% in the unmarked condition, this difference did not reach significance ($p = .08$).

In sum, the analyses of Comparison 1 revealed that children hearing the forms in their languages marked for false belief gen-

Table 5
Mean Number of Correct Responses on the Think *and* Look for *Questions by Age, Language, and Condition in Study 1*

Language and age	Condition: Marked	Unmarked	*Think* question second
Think question			
Turkish			
3-year-olds	1.63	1.73	—
4-year-olds	3.43	2.00	—
PR Spanish			
3-year-olds	1.50	2.00	—
4-year-olds	3.29	2.00	—
BR Portuguese			
3-year-olds	—	1.36	2.17
4-year-olds	—	2.80	3.07
English			
3-year-olds	—	1.33	1.55
4-year-olds	—	2.43	3.38
Look for question			
Turkish			
3-year-olds	1.75	2.09	—
4-year-olds	3.14	2.92	—
PR Spanish			
3-year-olds	1.00	2.67	—
4-year-olds	2.86	2.60	—
BR Portuguese			
3-year-olds	—	1.86	1.42
4-year-olds	—	2.60	3.14
English			
3-year-olds	—	1.44	2.11
4-year-olds	—	2.93	2.69

Note. Dashes indicate that no children were tested in those conditions. PR = Puerto Rican; BR = Brazilian.

erally answered more correctly to questions with the term *think* than did children without such forms in their languages. Both their mean scores and the number of children performing at a high level confirmed this finding. The lack of an interaction between age and condition revealed that this pattern of results tended to hold for both 3- and 4-year-olds. Importantly, there was no carryover from their high level of performance on that question to the more general *look for* question. Marking seemed to help the speakers of a marked language only locally, that is, to answer more correctly only those questions in which the actual marked form was used.

Comparison 2: Turkish and PR Spanish Two-Language Analyses

Here we again addressed the generality of the influence of marking for false belief in a language, but this time we compared the performances of speakers of marked languages who either heard those markings or heard the more neutral forms in the same languages. We also considered whether main verb marking in Turkish and clitic marking in PR Spanish had similar or different effects.

Because there were only 5 PR Spanish-speaking 3-year-olds remaining in the sample, we combined across ages for a 2 (languages: Turkish or PR Spanish) × 2 (condition: marked or unmarked) repeated measures ANOVA on uneven samples, on the *think* question and on the *look for* question. The analysis on the *think* question revealed a main effect of condition, $F(1, 58) = 6.14$, $p < .05$, but no effect of language and no interaction. The children hearing the marked forms of their languages answered more correctly ($M = 2.81$) than did children hearing the unmarked forms ($M = 1.90$). Seventy-one percent of children in the marked condition got 3–4 *think* questions correct, compared with only 32% of children in the unmarked condition, $\chi^2(1, N = 62) = 9.30$, $p < .01$. The analysis on the *look for* question yielded no significant effects.

These results show once again that marking for false belief in a language seems to have a local rather than a general effect. Moreover, the ANOVAs for the *think* and the *look for* questions did not reveal significant differences between main verb and clitic markings, which suggests that both kinds of markings may have similar local influences. However, given the small sample size and the consequent limited power, we cannot decisively conclude that differences in how false belief is explicitly marked do not influence false-belief understanding.

Comparison 3: BR Portuguese and English Order Analyses

Here we compared the performance of children who received *think* questions either before or after *look for* questions. Because in BR Portuguese the verb for *think* can also carry the meaning of *find*, BR Portuguese speakers may perform worse than English speakers on the *think/find* question, especially if hearing the *look for* question first primes a *find* interpretation. An interaction between language and order would support this hypothesis. A 2 (language: BR Portuguese or English) × 2 (order: *think* question first or *think* question second) × 2 (age: 3-year-olds or 4-year-olds) repeated measures ANOVA on uneven samples with answers to the *think* question as the dependent variable yielded a main effect of age, $F(1, 92) = 38.96$, $p < .01$, with 4-year-olds ($M = 2.91$) producing more correct answers than 3-year-olds ($M = 1.61$), and a main effect of order, $F(1, 92) = 7.15$, $p < .01$, with children hearing the *think* question second performing better ($M = 2.65$) than those hearing that question first ($M = 2.06$). There were no significant interaction effects. Only 37% of children hearing the *think* question first got 3–4 correct, whereas 63% of those hearing it second did, $\chi^2(1, N = 100) = 6.73$, $p < .05$. The results of the ANOVA with answers to the *look for* question as the dependent variable yielded only the usual significant main effect of age and no significant interactions.

In sum, both BR Portuguese and English speakers performed better on the *think* question when asked first about where the character would look for an object than when asked first about where the character thought the object was. Still, they did no better on the *look for* question itself depending on its order. The benefit to the second-place order may have accrued to the *think* question because the children had been engaged in the task by the earlier search question and then had time to consider the character's possible state of mind before having to answer a question about it. In any case, there was no indication that the polysemy of *achar* caused the BR Portuguese speakers to perform differently from the English speakers in either order.

Explanations

We coded all the responses to Question 3 (the request for explanations of experimenter or character behavior) given by the 118 children who participated in Comparison 1 whether or not they were later dropped from the statistical analyses. Some of the data from 1 PR Spanish speaker were lost, resulting in 469 total responses. Unsurprisingly, more than one third of the responses were irrelevant (171), and another 29% were coded as either action/perceptual or reality. Three- and 4-year-olds gave equal numbers of desire explanations (50 for each age group). Fewer than 20% of all responses (87) fell into the three coding categories (belief, lack of knowledge, or information seeking) that were most indicative of belief or knowledge understanding, and 70% of these were produced by 4-year-olds. Importantly, only 3 such explanations were forthcoming from Turkish- or PR Spanish-speaking 3-year-olds, whereas 23 came from BR Portuguese- and English-speaking younger children. Turkish- and PR Spanish-speaking 4-year-olds, with 20 such explanations, showed more understanding than their younger compatriots, although BR Portuguese- and English-speaking older children still did better, with 31. Thus, having a false belief term in one's language and hearing it in these scenarios did not encourage sophisticated responses to the requests for explanations of behavior.

Discussion

In this cross-linguistic study, we addressed the question of whether speakers of languages with explicit false belief markings would be advantaged on false belief tasks over children speaking languages without such markings. Our findings consistently point to a local advantage for speakers of marked languages. Turkish and PR Spanish speakers did better than speakers of BR Portuguese and English on false belief questions but only when the explicit false belief marker appeared in the question. They did no better on *look for* questions. They also did better on *think* questions, but not *look for* questions, than speakers of their own languages who heard a neutral rather than an explicit false belief term in the *think* question. How false belief was made explicit, whether by a main verb in Turkish or by the addition of a clitic in PR Spanish, did not seem to make a difference.

Interestingly, the advantage, local as it was, did have some influence even on 3-year-olds. Those without a false belief term in their language performed at levels worse than chance, whereas the performance of those with a false belief term at least was at chance. (See Wellman et al., 2001, for comparable poor performances of 3-year-olds.) Still, 4-year-olds generally performed better than 3-year-olds on all questions.

However, no local advantage of hearing the marked term extended to the Turkish- and PR Spanish-speaking children's explanations. The impressions of their experimenters were that, rather than being generally verbally reticent, Turkish- and PR Spanish-speaking children were reluctant to offer reasons for expected adult behavior to an adult experimenter. Thus, cultural differences may have affected the explanation data (see also Lillard, 1998).

One concern in this study was that although we had tried to control for socioeconomic differences among our language groups, the PR Spanish group seemed more heterogeneous than the others. To address this concern, we had eliminated the data of participants who either did not score at least 3 correct on the reality/control question or whose parents ranked below 50 on the SEI. The fact that there were no language differences in the analyses comparing the remaining PR Spanish speakers with the Turkish speakers suggests that we were successful in eliminating any extraneous differences that might have been attributed to SES and not type of explicit false belief marking.

Nonetheless, whether SES is a factor in the development of false belief understanding is an important question. As numerous studies have shown, SES makes a significant contribution to various aspects of both language experience and false belief understanding (e.g., Cutting & Dunn, 1999; Hart & Risley, 1995; Hoff-Ginsberg, 1998; Holmes, Black, & Miller, 1996). In order to address directly the role of SES (as measured by the SEI) in relation to differences in the explicit linguistic marking of false belief, we conducted a second study, collecting in advance data on parental occupation for two additional groups of PR Spanish and English speakers and examining anew whether the lexical explicitness of false belief marking would influence the performance on false belief tasks of groups differing in SES.

Study 2

Method

Participants

Twenty-three PR Spanish speakers, 12 whose parents ranked below 50 on the SEI (7 boys and 5 girls, mean age = 4 years 6 months) and 11 whose parents ranked 50 or above (6 boys and 5 girls; mean age = 4 years 7 months), participated in Study 2. In addition, 27 monolingual English speakers (almost all of them Caucasian), 14 whose parents' SEI scores were less than 50 (6 boys and 8 girls, mean age = 4 years 7 months) and 13 whose parents' scores were 50 or above (8 boys and 5 girls, mean age = 4 years 7 months), also participated. All participants were native speakers of their respective languages, and all attended either private or Head Start preschools.

Design and Procedure

We used a 2 (language: PR Spanish or English) $\times$ 2 (SES: $>$ 50 or $<$ 50 on the SEI) design and tested only 4-year-olds. All PR Spanish speakers heard the explicitly marked false belief form, and all speakers heard the *think* question before the *look for* question. The procedure was the same as in Study 1.

Results

Preliminary analyses revealed no significant differences in either age or SES (as measured by the SEI) between the two language groups. As in the earlier study, we eliminated from further statistical analyses all children who had scored less than 3 on the control/reality question. Six PR Spanish-speaking children, 5 in the low-SEI group and 1 in the high-SEI group, were eliminated, leaving a total of 17 PR Spanish speakers. No English-speaking children were eliminated.

Table 6 shows the mean correct responses to the *think* and *look for* questions for the two language groups by SES. We performed a two-way repeated measures ANOVA for unequal samples on the scores to the *think* questions and on the scores to the *look for* questions. The ANOVA on the *think* questions revealed that PR

Table 6
Mean Number of Correct Responses on Think *and* Look for *Questions by Language and Socioeconomic Status as Measured by SEI in Study 2*

	SEI group	
Language	High	Low
	Think question	
PR Spanish	3.90	1.29
English	3.17	.83
	Look for question	
PR Spanish	2.30	2.00
English	3.33	1.83

Note. SEI = Socioeconomic Index of Occupational Prestige; PR = Puerto Rican.

Spanish speakers produced significantly more correct responses to them than did English speakers, $F(1, 37) = 4.95$, $p < .05$, and that high-SEI speakers were significantly more often correct than were low-SEI speakers, $F(1, 37) = 86.18$, $p < .01$. There was no interaction. The ANOVA on responses to *look for* questions revealed no main effect of language and no interaction. However, there was a significant effect of SES: High-SEI children performed better than low-SEI children, $F(1, 37) = 5.38$, $p < .05$.

Analyses against chance on the *think* questions (chance = 2) revealed that, overall, PR Spanish speakers performed significantly better than expected by chance, $t(16) = 2.46$, $p < .05$, whereas English speakers did not ($p > .9$). The analyses further revealed that low-SEI children performed significantly worse than expected by chance, $t(18) = -4.94$, $p < .01$, whereas high-SEI children performed significantly better than expected by chance, $t(21) = 8.19$, $p < .01$.

Analyses of the number of children who answered the *think* questions correctly on at least 3 out of 4 tasks were consistent with the parametric analyses described above. Fifty-nine percent of the PR Spanish-speaking children got 3–4 *think* questions correct, whereas only 42% of the English-speaking children did so. This difference, however, did not reach significance ($p > .2$). Regarding SEI group differences, only one low-SEI child got 3–4 *think* questions correct, whereas 86% of the high-SEI children did so, $\chi^2(1, N = 41) = 26.84$, $p < .01$.

In sum, the children speaking the language with an explicit marking for false belief did better than the children speaking the unmarked language but only when they heard the explicit marking in the questions asked of them. Importantly, for speakers of both marked and unmarked languages, SES measured by the SEI had a major effect on the children's performance on false belief tasks regardless of how the false belief question was asked.

Discussion

Study 2 confirmed the findings of Study 1. Children in both studies, whether Turkish or PR Spanish speakers, who heard the false belief markers in a question performed better than children who did not. Study 2 also confirmed that this advantage for speakers of such languages appeared to be local rather than general, because no such advantage accrued to them when answering the *look for* question. In contrast, higher SES seemed to bestow a general advantage: It had a positive effect on both PR Spanish speakers' and English speakers' responses to both *think* and *look for* questions. Thus, in the sense that they may exert a general influence, socioeconomic considerations seem more central to the development of false belief understanding than does explicit lexical marking in one's language.

General Discussion

We began with the question of whether differences in the lexical explicitness with which languages express false belief might influence children's ability to pass false belief tasks. The answer we found by examining the question cross-linguistically was a limited "yes." Having an explicit term for false believing in their language facilitated performance on false belief tasks even for 3-year-olds (who nonetheless generally performed worse than 4-year-olds) but only when they had to respond to questions that used the explicit term.

Our finding adds importantly to the literature on cross-cultural research in several ways. First, it clarifies the finding of previous within-language research which suggested that for Chinese, the use of explicit terms facilitated correct responding on false belief tasks compared with nonexplicit terms (Lee et al., 1999). In Study 1, not only did we replicate that within-language finding with Turkish and PR Spanish speakers, but our design enabled us to show that the explicitness effect is local, that is, limited to *think* questions only and not *look for* questions. Also, we found that the particular linguistic form conveying the local influence seemed to be irrelevant: The PR Spanish clitic worked as well as the specific Turkish main verb. Thus, although as noted previously, this finding needs to be viewed with caution, the type of explicitness that exerts influence seems to be general, whereas the range of such influence is local.

Moreover, with the Study 1 design we examined languages with and without explicit terms, comparing BR Portuguese and English native speakers, who would never have even heard an explicit term in conversation, with Turkish and PR Spanish native speakers, who have them in their language. Again we found a local rather than a general effect of explicitness. Interestingly, we found a significant effect despite potential cultural differences in the countries, such as the willingness to talk about mental states, which might have had some impact on performance scores (see Lillard, 1998; Wellman et al., 2001). Indeed, we interpreted the limited explanation data as very possibly reflecting such cultural differences. Thus, we conclude that the local effect of lexical explicitness on the *think* question is a robust one.

One possible reason that cultural differences may have played relatively little role in our Study 1 participants' performances is that we had tried to minimize such differences by focusing on children in schools serving largely middle- to upper-middle-class communities. Experimenter impressions were that we were least successful in this effort in Puerto Rico. However, our strategy of eliminating not only children who answered fewer than 3 of the 4 control/reality questions correctly but also children of parents with low-ranking occupational prestige scores ensured that the remain-

ing PR Spanish-speaking sample in Study 1 was comparable to the other three samples.

In Study 2, our findings of effects of SES for both PR Spanish and English speakers are consistent with findings from previous studies showing such influences on aspects of false belief reasoning in English-speaking children (e.g., Cole & Mitchell, 1998; Cutting & Dunn, 1999; Holmes et al., 1996). Moreover, we found that both PR Spanish- and English-speaking children of high-SEI parents were significantly more likely to answer both the *think* and *look for* questions correctly than were children of low-SEI parents. Thus, SES seems to have an even broader effect on children's false belief understanding than does language explicitness, which affected only the *think* question responses in Study 1.

Interestingly, we found only a main effect of explicitness of marking and no interaction between it and SEI group on the *think* question, which suggests that even the low-SEI children were locally influenced by explicitness and that SES and explicitness of marking may be independent factors in children's developing false belief understanding (see also Cutting & Dunn, 1999). This finding further intimates that despite widespread differences in linguistic input associated with SES (Hoff-Ginsberg, 1998), false belief verb marking in PR Spanish is sufficiently available and salient to have some effect even on children with low-SEI parents. Nonetheless, the strength of the SEI group finding mandates controlling for SES when studying linguistic influences on false belief reasoning and, as others before us have argued, when investigating further the possible reasons for the frequently observed socioeconomic effects on understanding of mind.

The wider range in SES of the PR Spanish speakers in Study 1 does not fully account for their poor performance on the control/reality question that resulted in half of them being eliminated from statistical analyses. Even in Study 2, in which PR Spanish speakers performed better on the reality questions, 6 of them still had to be eliminated (compared with 0 English speakers). Without a definitive answer for why even PR Spanish-speaking children of parents with a high occupational prestige ranking might have had more difficulty with the reality questions than did speakers of other languages, we offer the following as a possible explanation. The reality question in PR Spanish may be more ambiguous than in the other languages, requiring a pause before "de verdad" in "Donde estan las crayolas, de verdad?" to arrive at the meaning "Where are the crayons, really?" (as opposed to "Where are the real crayons?"). Although the written PR Spanish protocol indicated the required pause with a comma and the experimenter was a native PR Spanish speaker, the experimenter nonetheless had the impression that some PR Spanish-speaking children may have misinterpreted the question—they would hesitate, peer at the picture of crayons on the (incorrect) box, and then point to it. Possibly the experimenter did not hesitate enough on all trials, or possibly not all young PR Spanish speakers are aware of the distinctions in meaning signaled by the pause. Future research should address this possibility. In any case, the likelihood that the reality questions were more ambiguous for PR Spanish speakers than for speakers of the other languages was confirmed by the fact that only the PR Spanish-speaking adult informant gave alternate readings of the reality question when asked to translate the English version into PR Spanish. (Adult informant translations revealed no other problems with either the reality question in the other languages or with the other questions in any language.)

One possible subtle influence of language on false belief understanding that we addressed was a potential effect of the polysemy of the BR Portuguese mental verb *achar*. We found that the mere fact of polysemy was not enough to affect BR Portuguese speakers' performance compared with English speakers' performance. Interestingly, we did find that the order of the questions asked had an effect on both groups of children. BR Portuguese and English speakers did better on the *think* question when it was asked after the *look for* question. This finding suggests that considering first where an actor will search for an object before one has to answer where that actor thinks it is generally facilitates correct responding to the *think* question. Although BR Portuguese- and English-speaking children hearing the *think* question second did not achieve the level of performance of Turkish- and PR Spanish-speaking children hearing explicitly marked *think* questions first (63% got 3 out of 4 correct compared with 71%, respectively), it would be useful to know whether the order of the questions improves the performance of even the children hearing explicitly marked language.

Our findings help us to clarify the role of language in the general cognitive developments related to understanding of mind. Several researchers have found relations between vocabulary size or syntactic skills and performance on false belief tasks, often using correlational analyses or comparing normal and special samples (see Astington & Jenkins, 1999; deVilliers, 2000). The finding of relationships between language and theory of mind warrants the hypothesis that speakers of different languages will show differences in their performance on false belief tasks. However, our results caution against accepting at this juncture a strong, causal relation between language and theory of mind. We found a relation between specific lexical markings of languages and better performance, but it was manifested only by what we have termed a local influence. Having an explicit lexical marking in their language is insufficient to make children generally better capable of understanding false belief. Hence, although we expect that other specific linguistic variables may be shown to have some influence on the development of understanding of mind, we are cautious in interpreting any found to date as strong causes. (For similar arguments regarding the limited role of specific language variables in conceptual development, see Heyman & Diesendruck, 2002; Papafragou, 2002; Sera, Bales, & del Castillo Pintado, 1997.)

Rather, we suggest that language may influence understanding of mind in a variety of ways—some direct, some indirect, some local, and some more general. We suspect that many of the previously reported relations between various language skills and false belief reasoning mostly reflect not specific, direct linguistic influences on such reasoning but rather more general, indirect influences of language use on the speeded cognitive development one finds in children from advantaged backgrounds (see Wellman et al., 2001, for a related argument about the role of language). Talk about others' feelings and thoughts, the recruitment of relatively sophisticated vocabulary to do so, and a generally enriched verbal environment very likely facilitate various cognitive skills, including false belief reasoning. One possibility, then, is that a social-linguistically enriched environment forms the basis for many of the relationships previously found between language and

false belief reasoning and between SES and false belief performance as well.[2]

However, enrichment of any sort, as a source of cognitive development, seems to have its limits. Although both parental SEI scores and explicit forms in a language can influence false belief performance, it is clear from our significant main effects of age (and no interactions) that even 3-year-olds with explicitly marked languages or from advantaged environments do not perform as well as their 4-year-old counterparts. Moreover, there are well-documented qualitative differences in the kinds of answers the younger children give to false belief questions. Thus, one can reasonably assume that the advantage of such enrichments is that they act like triggers for advancing to a higher stage of social thought rather than providing the basic materials out of which each child must create the higher level of thought himself or herself. The former approach still allows for variability in age of achievement in false belief reasoning while recognizing as well the ubiquity and relatively early acquisition of such reasoning around the globe.

In sum, our findings suggest a limited influence of explicit linguistic markedness for false belief on children's responses to false belief questions. More general are the influences of SES and of the cognitive level exemplified by age. Children of parents with high-status occupations and 4-year-olds show better performance on standard false belief tasks than do children of parents with low-status occupations and 3-year-olds. Explicit marking for false belief may best be seen as one characteristic of a language that by itself can exert a facilitating but limited influence on false belief reasoning.

[2] A full discussion of the complex relations and multidirectional influences among the various aspects of language, SES, and cognitive development is beyond the scope of this article. Nonetheless, reports of significant differences in the amount of parent–toddler linguistic interaction observed in different social classes (see Hoff, 2002, for a review) suggest that frequent early linguistic interaction may be an important feature of upper- and middle-class life, engendering a cascade of long-term repercussions for many facets of cognitive development, from vocabulary development to false belief reasoning to IQ scores on standardized tests. (See Shatz, 1994, on the development of social-linguistic intelligence in a middle-class child.)

References

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35,* 1311–1320.

Avis, J., & Harris, P. L. (1991). Belief–desire reasoning among Baka children: Evidence for a universal conception of mind. *Child Development, 62,* 460–467.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind.* New York: Oxford University Press.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77,* B25–B31.

Cole, K., & Mitchell, P. (1998). Family background in relation to deceptive ability and understanding of the mind. *Social Development, 7,* 181–197.

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development, 70,* 853–865.

deVilliers, J. G. (2000). Language and theory of mind: What is the developmental relationship? In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (pp. 83–123). London: Oxford University Press.

Dunn, J., Brown, J., & Beardsall, L. (1991). Family talk about feeling states and children's later understanding of others' emotions. *Developmental Psychology, 27,* 448–455.

Goetz, P. (2003). The effects of bilingualism on theory of mind development. *Bilingualism, 6,* 1–15.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Brookes Publishing.

Heyman, G. D., & Diesendruck, G. (2002). The Spanish *ser/estar* distinction in bilingual children's reasoning about human psychological characteristics. *Developmental Psychology, 38,* 407–417.

Hoff, E. (2002). Causes and consequences of SES-related differences in parent-to-child speech. In M. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 147–160). Mahwah, NJ: Erlbaum.

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics, 19,* 603–630.

Holmes, H. A., Black, C., & Miller, S. A. (1996). A cross-task comparison of false belief understanding in a Head Start population. *Journal of Experimental Child Psychology, 63,* 263–285.

Lee, K., Olson, D. R., & Torrance, N. (1999). Chinese children's understanding of false beliefs: The role of language. *Journal of Child Language, 26,* 1–21.

Lewis, C., & Mitchell, P. (Eds.). (1994). *Children's early understanding of mind: Origins and development.* Hove, England: Erlbaum.

Lillard, A. S. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin, 123,* 3–32.

Nakao, K., & Treas, J. (1992). *The 1989 Socioeconomic Index of Occupations: Construction from the 1989 Occupational Prestige scores* (General Social Survey Methodological Report No. 74). Chicago: University of Chicago, National Opinion Research Center.

Papafragou, A. (2002). Mindreading and verbal communication. *Mind and Language, 17,* 55–67.

Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA: MIT Press.

Sera, M. D., Bales, D. W., & del Castillo Pintado, J. (1997). *Ser* helps Spanish speakers identify "real" properties. *Child Development, 68,* 820–831.

Shatz, M. (1994). Theory of mind and the development of social-linguistic intelligence in early childhood. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 311–329). Hove, England: Erlbaum.

Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental terms: A systematic investigation of the first reference to mental state. *Cognition, 14,* 301–321.

Vinden, P. G. (1996). Junin Quechua children's understanding of mind. *Child Development, 67,* 1707–1716.

Wellman, H. M. (1990). *The child's theory of mind.* Cambridge, MA: MIT Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false-belief. *Child Development, 72,* 655–684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 53,* 45–57.

(*Appendix follows*)

Appendix

Description of the Languages

The Languages With Specific False Belief Terms

Turkish

Among the languages in this study, Turkish has the most elaborate system of encoding belief. This language parses thoughts and beliefs on a scale of certainty and distinguishes the ways in which information is acquired. An example of the latter is the evidential suffix *-mIş*, which is used to indicate second-hand information or information acquired by induction from apparent clues. To express thoughts and beliefs, the language uses several different verbs. The verb *düşün* is used to refer to mental activity such as thinking and to express a belief that is considered true or of unknown truth value, as in Example 1 below. The verbs *san* and *zannet* (verbs with similar meanings but with stylistic differences) are used to signal false belief. These verbs are more common than *düşün* in daily language. There are cases in which they can also imply uncertain beliefs, but only when used in the present tense, and even then they strongly emphasize the possibility of being mistaken. Thus, *san* in Example 2 implies a false belief.

1. "Can top-un nerede ol-dugu-nu [düşünüyor]?"
   Can ball-GEN where be-COMP-ACC [think-PROG]?[A1]
   Where does Can think the ball is?
2. "Can top-nerede [sanıyor]?"
   Can ball where [think-PROG]?
   Where does Can think (falsely) the ball is?

In addition to semantic differences, the verbs differ with regard to the structural complexity of the constructions in which they occur and their frequency of use in the language. The use of *san* is less complicated both syntactically and morphologically than *düşün.* The latter must have a complement clause in which the verb is marked with some form of the complementizer suffix *-DIK* and a possessive suffix agreeing with the subject, which is also marked with a genitive suffix, as in Example 1 above; but *san* can take a complement clause without any of the endings obligatory for *düşün*, as in Example 2. In addition, *düşün* is used in more formal contexts, whereas *san* is used more frequently both by children and their interlocutors.

Puerto Rican Spanish

In Puerto Rican Spanish, the system for encoding belief is somewhat different from that in Turkish. As in Example 3, the verb *pensar* is used to express mental activity and the possibility of an event. (These represent only a subset of the meanings of the Turkish *düşün.*) The most common verb to express belief is *creer*. Neutral with regard to the truth of the belief, it can be used with true beliefs or with beliefs of unknown truth status, as in Example 4. To distinguish false beliefs from true or unknown truth-value belief states, the reflexive clitic *se* is added to the verb *creer*, resulting in the form *creer-se* illustrated in Example 5.

3. "Estoy [pensando] ir al cine."
   (I) am [thinking] of going to the movies.
   I am thinking of going to the movies.
4. "Maria [cree] que su libro esta debajo de la mesa y, de hecho, ahi esta."[A2]
   Maria [believes] that her book is under the table and, in fact, it's there.
   Maria thinks that her book is under the table, and in fact, it's there.
5. "Maria [se cree] que su libro esta debajo de la mesa, la verdad es que esta en la gaveta."
   Maria [believes (falsely)] that her book is under the table, but the truth is, it's in the drawer.
   Maria thinks her book is under the table, but the truth is, it's in the drawer.

Despite the semantic differences among these verbs, they share some syntactic similarities. All three verbs, *pensar*, *creer*, and *creer-se*, occur in constructions either taking a complement clause introduced by the complementizer *que* ("that") or with an infinitive verb. *Creer* and *creer-se* differ syntactically from *pensar* in that they can also occur in constructions preceding nominals that function as direct objects, as in Example 6, confirming the fact that *pensar* refers to the act of thinking.

6. "Ella [cree] al cura."
   She [believes] the priest.

In addition, *creer-se* may be used to express false self-attributions, in which case the verb takes a qualifier, as in Example 7.

7. "Maria [se cree] tan inteligente."
   Maria [believes (falsely)] so intelligent.
   Maria thinks that she is so intelligent.

Thus, although *creer* and *creer-se* differ morphologically by the addition of the clitic to express a known false belief or a false self-attribution, they are very similar syntactically.

The Languages Lacking Specific False Belief Terms

English

In English, as with other languages, the semantic space of thought and belief is partitioned with several verb forms (e.g., *think* and *believe*). Like the Spanish *pensar*, *think* can refer to a mental activity ("I'm thinking about class this afternoon"), the possibility of an event ("I think it will snow later"), or the planning of an activity ("I'm thinking of going to the movies"). However, of the verbs in the four languages in our study, the English *think* is the most polysemous, being used to express a belief regardless of its truth value, as in Example 8.

8. "John [thinks] we went to Gratzi last night, but we went to the Earle."

The verb *believe* can also be used to express belief regardless of its truth value, but *think* is more common in everyday English, especially in children's speech (Bartsch & Wellman, 1995; Shatz, Wellman, & Silber, 1983).

[A1] GEN (genitive), COMP (complementizer), ACC (accusative), and PROG (progressive) represent the inflectional suffixes denoting grammatical relations in a highly inflected language like Turkish. Thus, PROG refers to the progressive tense marking on the verb.

[A2] Note that *que* in Puerto Rican Spanish and Brazilian Portuguese is a complementizer in Examples 4, 5, 10, and 11. In the English Example 8, the complementizer *that* is understood but deleted from the surface structure.

Brazilian Portuguese

As in Puerto Rican Spanish, Brazilian Portuguese also uses the verb *pensar* to express mental activity or the possibility of an event, as in Example 9.

9. "Estou [pensando] em ir ao cinema."
   I'm [thinking] of going to the movies.

The verb used to talk about beliefs is *achar*, which also means to find. As in English, Brazilian Portuguese uses the same linguistic form to express both belief (Example 10) and false belief (Example 11), but it differs from English in that *achar* has the additional factive meaning (Example 12).

10. "Maria [acha] que seu livro esta embaixo da mesa, e realmente esta la."
    Maria [thinks] that her book is under the table, and in fact, it's there.
11. "Maria [acha] que seu livro esta embaixo da mesa, mas na verdade, esta na gaveta."
    Maria [thinks] that her book is under the table, but the truth is, it's in the drawer.
12. "Maria [achou] seu livro embaixo da mesa."
    Maria [found] her book under the table.

In Brazilian Portuguese, then, *achar* is polysemous in that it can refer to two different states, a mental state and a factive state. Although *find* in English can be used to refer to a mental state, as in Example 13, it cannot carry the notion of false belief, as in Example 11.

13. "I find your statement to be offensive."

Different syntactic behaviors are associated with the different meanings of *achar*. For example, when it is used to refer to a mental state, it usually takes a complement clause starting with the complementizer *que* (Examples 10 and 11). When used with the factive meaning, *achar* takes a direct object complement (Example 12). But often cases are indistinguishable syntactically and meanings must be disambiguated via context and pragmatics, as in Example 14.

14. "Eu nao [achei]."
    I didn't [think] (anything).
    or
    I didn't [find] (anything).

Received August 27, 2002
Revision received January 29, 2003
Accepted February 4, 2003 ■

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor.* High School Teacher and International Affiliates receive subscriptions to the *APA Monitor,* and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association,* and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.