British Journal of Developmental Psychology (2005), 23, 451–469

The British Psychological Society

www.bpsjournals.co.uk

Guessing versus choosing – and seeing versus believing – in false belief tasks

Heidemarie Lohmann, Malinda Carpenter* and Josep Call
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Three- and 4-year-old children were tested using videos of puppets in various versions of a theory of mind change-of-location situation, in order to answer several questions about what children are doing when they pass false belief tests. To investigate whether children were guessing or confidently choosing their answer to the test question, a condition in which children were forced to guess was also included, and measures of uncertainty were compared across conditions. To investigate whether children were using simpler strategies than an understanding of false belief to pass the test, we teased apart the seeing-knowing confound in the traditional change-of-location task. We also investigated relations between children's performance on true and false belief tests. Results indicated that children appeared to be deliberately choosing, not guessing, in the false belief tasks. Children performed just as well whether the protagonist gained information about the object visually or verbally, indicating that children were not using a simple rule based on seeing to predict the protagonist's behaviour. A true belief condition was significantly easier for children than a false belief condition as long as it was of low processing demands. Children's success rate on the different versions of the standard false belief task was influenced by factors such as processing demands of the stories and the child's verbal abilities.

After hundreds of studies of children's theory of mind and false belief understanding, many questions still remain about what exactly children are doing when they pass false belief tests. The most important of these questions is, naturally: do children need an understanding of others' false beliefs to pass the tests, or can they be passed in some other way? In the current study, we investigate children's strategies and patterns of responding during various false belief tests as compared with other similar tests, to determine whether children are truly passing the tests using an understanding of false belief. There are two important issues: the first involves interpretations of the pattern of results across children, and the second involves investigating by what means these children achieve correct scores in these tests.

**Correspondence should be addressed to Malinda Carpenter, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany (e-mail: carpenter@eva.mpg.de).*

DOI:10.1348/026151005X26877

First, in a meta-analysis of 178 studies, Wellman, Cross, and Watson (2001) found that children younger than 3.5 years typically fail traditional false belief tests such as the 'Sally-Anne' test (Baron-Cohen, Leslie, & Frith, 1985), children older than 4 years pass these tests, and children between 3.5 and 4 years perform at chance level as a group. This finding of chance-level responding for the middle age group could be produced in two ways, at the extremes: either all children guessed randomly, or half answered incorrectly and half answered correctly. In a recent study, Carpenter, Call, and Tomasello (2002) used a new procedure to investigate whether young 3-year-olds could pass a simplified version of a false belief test, and if so, whether they did this by guessing or confidently choosing. They presented children with a change-of-location task during a natural interaction. To determine whether children were guessing or choosing, during the false belief test they measured children's latency to choose a location and signs of uncertainty. They compared these measures with those during two true belief tests and a version of the test in which children were forced to guess because they were not given enough information. Carpenter *et al.* found that even these young 3-year-olds were not guessing during the false belief test. Instead, they were confidently choosing a response – many of them a correct one.

Ruffman, Garnham, Import, and Connolly (2001) also studied children's uncertainty during false belief tests, using an anticipatory looking procedure like that of Clements and Perner (1994), and asking children to bet counters corresponding to how certain they were of their answer. Whereas Carpenter *et al.* (2002) found no differences in uncertainty between 3-year-old passers and failers of their test, Ruffman *et al.* found an interesting developmental pattern of increasing and decreasing uncertainty: younger failers were certain of their answers, older failers and younger passers were more uncertain, and older passers were again certain. The difference in findings between the two studies may be due to differences in the explicitness of the studies' uncertainty measures as well as to differences in the age range of children participating.

One goal of the current study was to investigate children's levels of uncertainty using a more traditional false belief task. We tested children in the 'chance' age range (i.e. $3\frac{1}{2}$–4 years) using videos of puppets in standard change-of-location false belief situations (using variations of the 'Sally-Anne' task of Baron-Cohen *et al.*, 1985). We also included a condition in which children were forced to guess when answering the test question, as well as other types of true and false belief conditions (see below), and we collected latency and uncertainty measures in all conditions to determine whether children were guessing or confidently choosing their answer.

Second, Carpenter *et al.* (2002) found that some young 3-year-olds appeared to be truly passing their false belief test, and other studies (e.g. Clements & Perner, 1994; Ruffman *et al.*, 2001) have found more implicit evidence of some false belief understanding in old 2-year-olds and young 3-year-olds. This raises the question of how these young children were passing and what factors allowed for a successful performance. For example, did they use an understanding of false belief or some other, simpler understanding? In traditional false belief tasks (e.g. Baron-Cohen *et al.*, 1985), children watch a protagonist place an object in one location and leave. The object is then moved by another actor to a new location. When the protagonist returns, children are asked where the protagonist will look for the object. If children correctly answer the original location, this is taken as evidence that they understand that the protagonist holds a false belief about the location of the object. However, there are other, simpler possible reasoning strategies that would also result in correct performance on this task.

For example, some authors (e.g. Garnham & Ruffman, 2001; Robinson & Mitchell, 1992; Wimmer & Perner, 1983; and others – see Garnham & Ruffman, 2001, for a review) have proposed that young children could use a seeing = knowing rule. These authors tested whether children use this rule by adding a third location to the test. They reasoned that if children assumed that people who have not seen something always do the wrong thing, then children should choose the third, irrelevant location equally as often as the other wrong (from the point of view of the story protagonist) location (i.e. where the object originally was before being moved). They found that children rarely choose the irrelevant location, suggesting that children were using more than a simple not-seeing = not-knowing rule.

In a similar vein, it is also possible to respond correctly to false belief tests by using the simple rule that people tend to look for objects where they last saw them. By using this rule, children could answer correctly without having any understanding of belief, or even knowledge.[1] Perner, Leekam, and Wimmer (1987) tested a similar hypothesis but did so in a complicated way involving complex verbal information concerning instructing, anticipating, and forgetting, and no change of location. They concluded that although most children understood false belief by the age of 4 years, some correct responses of the younger children might have been based on a physical association strategy. In the current study, we attempted to determine whether children had any understanding of the protagonist's knowledge (or belief) by teasing apart the seeing-knowing confound inherent in the task in a different way. We tested whether children could update the protagonist's knowledge using a different modality – not seeing but hearing. That is, we added conditions in which the protagonist puts an object in one location and then upon his return is told (correctly or incorrectly) where the object now is. If children used a simple rule of people tend to look for objects where they last saw them, it would lead them to an incorrect answer in these conditions.

We are unaware of any previous studies that have used a telling condition in a false belief task, although several studies have investigated children's use of verbal information to update their own and others' knowledge. For example, Wimmer, Hogrefe, and Perner (1988; Experiment 1) had 3- to 5-year-old children judge other children's informational status after they were or were not informed either visually or verbally about the contents of a box. Almost all the 3-year-old children failed this task, and only about half of the 4-year-olds passed. Children found the telling condition equally as easy (or difficult) as the seeing condition. Further studies with easier test questions have found better performance by 3-year-olds but have only investigated children's understanding of their own informational status (e.g. Gopnik & Graf, 1988; O'Neill & Gopnik, 1991; Pratt & Bryant, 1990) or did not include telling conditions (Pillow, 1989, Pillow, 1993; Pratt & Bryant, 1990). It is thus an open question whether 3-year-olds understand that other people can update their knowledge by being told of an event that they have not seen.

While positive results in such telling conditions would provide evidence for an understanding of what factors influence the informational states of others, they would not answer the question of whether children need to have a representational

---

[1] *Note that a similar explanation is possible for 'deceptive box' tasks such as the Smarties test (Perner, Frith, Leslie, & Leekam, 1989). In this case, children could respond correctly based on what they know that people have seen in the past – they have seen Smarties, not pencils, in Smarties boxes. On the other hand, if children do pass deceptive box tasks using an understanding of false beliefs, this would explain why results on deceptive box and change of location tests do not always correlate with each other (e.g. Naito, 2003).*

understanding of belief states when passing such tests. According to Perner's (1991) view that beliefs have to be understood as representations before children master a false belief task, it is the falseness of the belief that makes the standard tasks difficult (see also Gopnik & Wellman, 1994; Wellman, 1990; Zaitchik, 1990). In line with this argument for a conceptual development, some studies have shown that preschool children pass true belief versions of the task earlier than comparable false belief versions of it (Wimmer & Perner, 1983). In contrast, Roth & Leslie (1998) argue that the difficulty of belief tasks does not stem from the falseness of the attributed belief, but from the demands of correctly calculating an attributed belief content. These authors thus assume a limitation in processing abilities which constrains children's performance in these tests. Some studies have supported this alternative view by finding no difference in passing rate between true and false versions of a standard belief task scenario (Roth & Leslie, 1998; Siegal & Beattie, 1991; Surian & Leslie, 1999).

In the current study, along with the standard false belief test, we included two different versions of a standard true belief test in order to explore processing influences: one in which there is no change of location and one in which the protagonist witnesses the change of location. We also included two versions of a telling condition in order to explore conceptual differences: one in which the object is moved and the protagonist is told this and one in which the object is not moved but the protagonist is told that it was. We included the last condition to gain further information about children's understanding of lies. Johnson (1997) has shown that as soon as 4-year-old children were able to understand a speaker's false belief, most could correctly judge the well-meaning intention of a person who holds a false belief and thus distinguish mistakes from lies. For younger children, Johnson and Maratsos (1977) and Zaitchik (1991) showed that $3\frac{1}{2}$-year-olds could correctly predict the future action of a protagonist who holds a false belief due to someone telling him a lie, but only when the child's own true belief was acquired by a less reliable source of information (by testimony). Thus, there is some evidence that under some circumstances (e.g. when the salience of the real location of the object is reduced), 3-year-old children appreciate the informational status of a protagonist in a false belief setting. However, children still failed in a standard false belief story setting, where the children witnessed the true location of the object. Thus, it is necessary to compare in a standard belief scenario different modes of information (visual vs. verbal) as well as the truth value of the belief (false vs. true belief) in 3-year-old children in order to tease apart the seeing-knowing confound and processing from conceptual factors in the traditional test.

Method

Participants

Eighty-five German children between $3\frac{1}{2}$ and 4 years of age participated in this study (mean age = 44 months, 14 days; *SD* = 2.06 months; range = 41 months, 10 days to 48 months, 30 days). There were 41 girls and 44 boys. Children were recruited from day-care centres in the city of Leipzig, Germany.

Procedure

Children were tested individually in a quiet room at their day-care centre. Following a brief warm-up period with unrelated toys, each child was shown six short videos (approximately 2.5 minutes each) and was asked questions at the end of each video. Each video depicted a different pair of puppets (e.g. two girls, two boys, a rabbit and a hedgehog, and a boy and a polar bear) who acted on and talked about an object (e.g. a necklace, a bicycle, a carrot, and a ball, respectively), moving it from one closed location to another (e.g. a drawer and a box, a shed and a garage, an upside-down cup and a box, and a basket and an upside-down cup, respectively). Figure 1 presents a photograph taken from one of the videos; see the Appendix for a script of each of the conditions in one of the story lines.

Each video consisted of a different modification of the standard false belief change-of-location story (Wimmer & Perner, 1983), in which a protagonist puts an object in location A, and then the object is moved to location B in the absence of the protagonist. Each video thus began with the first protagonist placing an object in location A (this location was counterbalanced for side of the video-screen for each condition). The six modifications of the standard false belief change-of-location story varied over the final location of the object (A or B) and, more importantly, over whether and how the main protagonist learned about the object's move (i.e. visually or verbally).

There were three true belief conditions, in which the main protagonist knew the correct location of the object at the time of the child's test questions. In each of these conditions, this came about in a different way: either the object remained in the original location or the protagonist learned about the change of location either visually (i.e. the protagonist witnessed the change of location) or verbally (i.e. the protagonist was told about the change of location), as follows:

No switch condition: After placing the object in location A, the main protagonist left. During the main protagonist's absence, the second protagonist appeared, discovered the object in location A, replaced the object in location A, and left.

Seen switch condition: After placing the object in location A, the main protagonist stayed and watched as the second protagonist moved the object to location B and left.

Tell condition: After placing the object in location A, the main protagonist left. During the main protagonist's absence, the second protagonist appeared, discovered the object in location A, and moved the object to location B. The main protagonist then

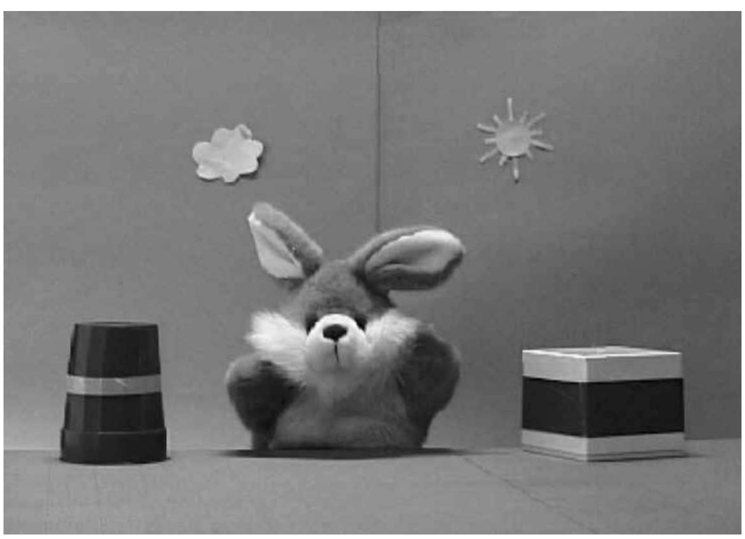

Figure 1. A photograph taken from one of the videos.

reappeared, and the second protagonist told the main protagonist that he/she had moved the object, and that it was now in location B. The second protagonist then left.

There were also two false belief conditions, in which the main protagonist had a false belief about the correct location of the object at the time of the child's test questions. Again, in each of these conditions, this came about in a different way: the protagonist's false belief was brought about either visually (i.e. the protagonist did not witness the object being moved to a different location) or verbally (i.e. the object was not moved but the protagonist was told that it was moved), as follows:

Unseen switch condition (i.e. traditional false belief scenario): After placing the object in location A, the main protagonist left. During the main protagonist's absence, the second protagonist appeared, discovered the object in location A, and moved it to location B. Then the second protagonist left.

Lie condition: After placing the object in location A, the main protagonist left. During the main protagonist's absence, the second protagonist appeared, discovered the object in location A, and replaced the object in location A. The main protagonist then reappeared, and the second protagonist told the main protagonist that he/she had moved the object, and that it was now in location B (even though he/she had not moved the object). The second protagonist then left.

In all five of these conditions children knew where the object really was at all times, including during the test questions. In order to measure how children responded when they were guessing, we included a final condition in which we forced children to guess during the test (and reality control) questions, as follows:

Guess control condition: The second protagonist appeared alone and introduced him/herself as a friend of the main protagonist. He/she mentioned that the object belonged to the main protagonist, who wanted to come and get it later, and then announced that he/she now was going to put the object away. Before he/she could do this, a large bus drove by and covered the entire scene on the video. When the scene was visible again, the second protagonist stated that he/she had just put the object into one of the two locations, either A or B. Then the second protagonist left and the main protagonist appeared. Thus, neither the main protagonist nor the child knew where the object was located at the time of the test question.

At the end of each of these conditions, when the second protagonist had left and the main protagonist had (re)appeared, the scene was frozen with the main protagonist standing alone and in the centre of the two closed locations, looking straight ahead. While this image remained on the screen, children were asked a test question: 'Where will (name of main protagonist) look for the (object) first?' This was followed by memory and reality control questions: 'Where did (name of main protagonist) put the (object) in the beginning?', and 'Where is the (object) now, really?', before continuing to the next story. Children could answer the questions either verbally or by pointing. Only uninformative feedback (e.g. 'mmm') was given.

Children watched all six stories in succession in a single testing session. Six different story plots with five conditions of varied starting locations (plus the guess condition) were possible, which would have led to a total of 66 short videos. Because there were so many possible combinations of order and story plot, we chose six different combinations (36 videos), with the provision that each story plot and condition were in each order position once and the start and final locations of the object were counterbalanced. Children were randomly assigned to one of these six combinations.

Upon completion of all six videos and their test questions, we also assessed children's language abilities using the vocabulary subset from the Kaufman Assessment

Battery for Children (Kaufman & Kaufman, 1994, German version) and the imitation of grammatical structures subtest from the Heidelberger Sprachentwicklungstest (Grimm & Schoeler, 1991).

Coding

All sessions were videotaped. From the videotapes, we scored children's answers to the test and control questions, their latency to answer the test question, and signs of uncertainty during children's answers to the test question. The latency (in seconds) to choose a location was defined from the moment the experimenter had finished her question until the child's final answer (location word or point). Uncertainty was scored on a yes/no basis. Signs of uncertainty included uncertain looks before and during the answer (e.g. to the experimenter), verbal markers (e.g. 'She'll look either here or here' or 'maybe'), or other non-verbal signs of uncertainty. Such other non-verbal signs of uncertainty included very slow speaking, speaking pauses, using a questioning tone, shrugging the shoulders, and starting to point but then withdrawing the hand. To assess inter-observer reliability, 17 of the 84 children (20%) representing 102 trials were chosen randomly and coded independently by two coders, blind to the story condition. There was good inter-observer reliability for the latency to choose ($r = .98$, $p < .001$) and the uncertainty measure (Cohen's $\kappa = 0.73$).

Results

In each condition, only children who passed both the reality and the memory control questions were included in the analyses. This resulted in different numbers of children in each condition (see Fig 2 for *N*s). We analysed the effect of the order of presentation of each condition to investigate whether children's answers in a given condition changed as a function of whether children had experienced other conditions first. The percentage of correct responses of those children who received a given condition in the first order[2] was compared to that of children who received it in all later orders. There were no significant differences between these two groups of children for any condition (Fisher tests: $p > .16$ in all cases). Consequently, order is not considered in any further analyses.

Children's answers to the test questions

Figure 2 presents the percentage of children who answered the test question (and both control questions) correctly for each condition. The majority of children (63–80%) passed each of the three true belief conditions and approximately 40% of the children passed each of the two false belief conditions. As expected, in the guess control condition, children chose each location equally often ($\chi^2 = .352$, $df = 1$, $p = .55$).

Often, studies of children's understanding of false belief compare children's responses to a chance value of 50%. The responses of children in the current study were significantly above chance (Binomial test with 50% as the expected value, one-tailed) in the no switch ($p < .001$) and tell conditions ($p < .04$), and approached significance in the seen switch condition ($p = .064$). Of particular interest is the comparison of the false belief conditions (unseen switch and lie) to chance. In these two conditions, in

[2] *These percentages were 86, 80, 60, 60, and 43%, for the no switch, seen switch, tell, unseen switch, and lie conditions, respectively.*

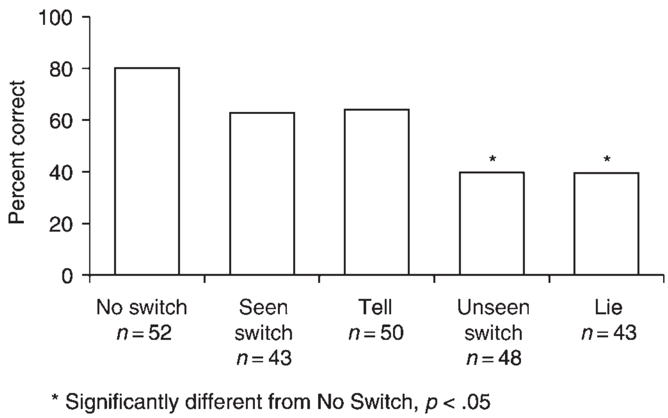

Figure 2. The percentage of children who answered the test question (and both control questions) correctly for each condition.

contrast, children responded at chance levels ($p = .19$ for unseen switch and $p = .22$ for lie), thus replicating the findings of the Wellman *et al.* (2001) meta-analysis for this age range.

However, the comparison to 50% chance can be misleading, for two reasons (Carpenter *et al.*, 2002). First, if children do not understand others' false beliefs, they will likely treat the protagonist as having a true belief. If this is the case, they should respond identically (i.e. in the current study, they should choose the same location) in the true belief and the false belief conditions. Second, because the two locations are not equally salient to children at the time of the test question (because the object is in one of them), the true chance value might not be 50% because children could be drawn to the real location of the object. To better assess children's false belief understanding, we thus compared what children did in each of the false belief conditions to what they did in the corresponding true belief conditions as a baseline. That is, we compared children's performance in the false belief conditions not to chance (50%) but rather to what they would do if they assumed that the protagonist had a true belief about the location of the object.

We used the percentage of incorrect responses in the appropriate true belief condition (i.e. seen switch for unseen switch, and no switch for lie) to estimate the percentage of correct responses that would be expected in the false belief conditions. In other words, if 37% of the children chose the wrong container in the seen switch condition, it was expected that (approximately) 37% of the correct responses in the unseen switch condition were achieved by a strategy different from an appreciation of false belief (i.e. false positives). Any positive deviation from these values would indicate a true appreciation of false belief or lying. Using this correction procedure revealed that children significantly exceeded the percentage of correct responses that would be expected from the use of a true belief strategy in the lie condition (40% observed vs. 20% expected; Binomial test, $p < .001$) but not the unseen switch condition (40% observed vs. 37% expected; Binomial test, $p = .42$). When the unseen switch condition was compared with a different true belief condition (no switch), however, children significantly exceeded the percentage of correct responses that would be expected from the use of a true belief strategy (40% observed vs. 20% expected; Binomial test, $p < .001$).

As a further test of whether children were truly passing the false belief tests, we investigated whether the group results also held for individual children by looking at patterns of performance between the matching true belief and false belief conditions, to see whether individual children could switch strategies and choose different locations across conditions. Table 1 shows that 6 of the 14 (43%) passers of the unseen switch condition also passed the corresponding seen switch true belief condition, and 8 of the 12 (67%) passers of the lie condition also passed the corresponding no switch true belief condition. Thus, roughly half of the children in this age range who passed the false belief test question (and the control questions) were likely passing the test with an understanding of false belief.

Table 1. Individual patterns for matching true belief and false belief conditions

	Seen switch			No switch	
Unseen switch	Pass	Fail	Lie	Pass	Fail
Pass	6	8	Pass	8	4
Fail	17	6	Fail	21	1

Other comparisons
Comparisons across conditions indicated no significant differences among the three true belief conditions (McNemar test: no switch-tell: $p = .11$, no switch-seen switch: $p = .69$ and tell-seen switch: $p = 1.00$) and no significant difference between the two false belief conditions (McNemar test: unseen switch-lie: $p = 1.0$).

Two comparisons were of particular interest. First, we compared the visual and verbal versions of the true belief and false belief tasks in order to answer the question of whether children could be passing the false belief tasks with an understanding of seeing rather than belief. There was no significant difference between the visual and verbal versions of the task: for the true belief conditions, seen and tell switch conditions, McNemar test, $p = 1.00$; for the false belief conditions, unseen switch and lie, McNemar test, $p = 1.00$. Thus, in the traditional false belief test situation (as well as in a true belief situation), children did equally well when they had to update the protagonist's beliefs with a modality other than seeing, thus undermining the idea that children could be using the simpler rule of 'the protagonist goes where he last saw the object' to solve the test.

Second, we compared the true and false belief tests in order to address different hypotheses about whether false belief tests should be more or equally difficult as compared with true belief tests. Children performed significantly better in the no switch compared to the unseen switch (McNemar test: $p = .027$) and the lie (McNemar test: $p = .001$) conditions. However, there were no significant differences in performance between the seen switch condition and the unseen switch (McNemar test: $p = .11$) or lie (McNemar test: $p = .21$) conditions. Thus, when task demands were increasingly complex (i.e. there was a switch or the mention of a switch), the differences between true and false belief conditions vanished for this age group. In order to assess the level of influence of processing demands on children's performance we compared conditions of similar complexity with each other. If processing demands are crucial, then we should find similar correlations between true and false belief conditions and within true or false belief conditions.

Table 2 shows the number of children who passed (and failed) each of the three true belief conditions. There were significant positive relations between the scores on these tasks ($r_s \geq .39$, $p \leq .014$, $N \geq 34$ in each case). Most children (78%) either passed or failed two or more true belief tasks, perhaps indicating the operation of a common mechanism across tasks. Note, however, that this result may be partly a result of the high percentage of children who passed both tasks. Table 2 also presents the number of children who passed (and failed) each of the two false belief conditions. There was no significant relation between these two variables ($r_s = .18$, $p = .42$, $N = 22$). Only 59% of the children either passed or failed both tasks. This suggests that children's success in these conditions did not depend mainly on one common factor. Finally, Table 2 shows those correlations between true and false belief conditions with similar processing

Table 2. The number of children who passed and failed each pair of true belief and each pair of false belief conditions and true and false belief conditions with similar processing demands (switch)

True belief conditions

	No switch	
Tell	Pass	Fail
Pass	23	2
Fail	8	6

Correlation: $r = .41$, $p = .009$, $N = 39$

	No switch	
Seen switch	Pass	Fail
Pass	22	2
Fail	4	6

Correlation: $r = .55$, $p = .001$, $N = 34$

	Tell	
Seen switch	Pass	Fail
Pass	21	5
Fail	6	8

Correlation: $r = .39$, $p = .0014$, $N = 40$

False belief conditions

	Unseen switch	
Lie	Pass	Fail
Pass	6	5
Fail	4	7

Correlation: $r = .18$, $p = .42$, $N = 22$

True and false belief conditions of similar processing demands

Seen switch	Unseen switch: Pass	Unseen switch: Fail
Pass	6	17
Fail	8	6

Correlation: $r = -.31$, $p = .06$, $N = 37$

Tell	Unseen switch: Pass	Unseen switch: Fail
Pass	6	20
Fail	11	5

Correlation: $r = -.45$, $p = .01$, $N = 42$

demands. If processing demands rather than the type of belief determined the outcome of each task, we would expect positive correlations between tasks with similar demands. Contrary to this expectation, there were only negative or non-significant relations. This means that processing demands alone cannot explain our results.

Finally, we investigated what factors might have influenced children's success or failure in the different conditions. Table 3 shows the results of a logistic regression assessing the influence of gender, age, vocabulary scores, and grammar scores on children's performance in each of the different conditions. This analysis is necessary because it allows us to study the effect of all factors simultaneously and assess which factors have a greater explanatory power. To assess the effect of vocabulary and grammar skills, we used a median split of children's verbal scores to compare children with high versus low scores in vocabulary or grammar. Coefficients below 1 indicate a negative relation between the factor and the dependent measure whereas coefficients above 1 indicate a positive relation. Only those coefficients that were significant are presented.

Table 3. Results of the logistic regression testing the influence of age, gender, vocabulary, and grammar scores in each of the conditions. Shown are the exponential coefficients (e^b – odds ratio), their associated significance, and the standardized effects (z values). Only those factors that produced a significant model are presented

	No switch ($N = 52$)	Seen switch ($N = 43$)	Tell ($N = 50$)	Unseen switch ($N = 48$)	Lie ($N = 43$)
Age			$-.56$ ($p = .00$) $z = 3.04$		
Gender					5.29 ($p = .03$) $z = 2.16$
Vocabulary				1.97 ($p = .00$) $z = 2.80$	1.38 ($p = .04$) $z = 2.02$
Grammar				1.38 ($p = .01$) $z = 2.57$	

Language measures had a significant effect on children's performance in both the unseen switch and lie conditions. In particular, children with high vocabulary and grammar scores performed better in the unseen switch task than those with lower scores in those language measures. Similarly, children with higher scores in vocabulary size also performed better in the lie condition than those with lower scores. Of the children with high vocabulary scores, 64% and 54% passed the lie and the unseen switch tests, respectively, whereas the number of successful children decreased to 29% and 26%, respectively in each of those conditions for children with low scores in the vocabulary test. An additional influence on performance in the lie condition was found for the gender of the child, with boys scoring significantly higher in the lie condition than girls.

In contrast, language measures or gender had no positive effect on children's performance in any of the three true belief conditions. The only significant model for a true belief condition revealed the effect of age being negatively related to performance in the tell condition, with younger children performing better than older ones.

Uncertainty measures

Previously we noted that children's scores were at chance level (50%) for all conditions except for the no switch and tell conditions. Chance performance is not easy to interpret: children may have been guessing or they may have been deliberately choosing their answer (half of them correctly and half of them incorrectly). In order to distinguish children who were guessing from those who were using a distinct strategy, we investigated how children behaved when they were forced to guess and compared it to their behaviour during those conditions in which there was no need to guess.

Figure 3 presents the mean latency to choose one of the alternatives in each of the experimental conditions compared to the guess condition. Children took significantly longer to answer in the guess condition compared to the tell ($t = 1.90$, $p = .032$) and the seen switch ($t = 2.699$, $p = .005$) conditions. Children also tended to take longer to answer in the guess condition compared to the unseen switch condition, but this difference was not significant ($t = 1.375$, $p = .09$). There were no significant differences between the guess condition and the no switch ($t = .936$, $p = .18$) or the lie conditions ($t = .726$, $p = .24$).

Table 4 presents the percentage of children who showed signs of uncertainty in each of the experimental conditions compared to the guess condition for three indicators of uncertainty. There were no significant differences between the guess and other conditions with regard to uncertain looks (McNemar tests: $p > .05$ in all cases). In contrast, children produced significantly more uncertain words in the guess condition than in the unseen switch and the seen switch conditions (McNemar tests: $p = .013$ in both cases). Similarly, children showed more non-verbal signs of uncertainty in the guess condition compared to the other conditions, in all but the no switch condition (McNemar tests: $p < .05$ in these cases). Approximately 27% of the children showed non-verbal signs of uncertainty in the guess condition compared to 14% or less in all other conditions. Moreover, combining these three uncertainty measures into an overall score revealed that the guess condition produced significantly more uncertainty than all the other conditions (Wilcoxon tests: $p < .02$ in all cases). Children who passed were not more or less uncertain than those who failed the test question in each of the experimental conditions (Fisher's tests, $p > .18$ in all cases), including the two false belief conditions. To summarize, children behaved differently in all conditions as compared with the guess condition, by taking longer to respond or showing more

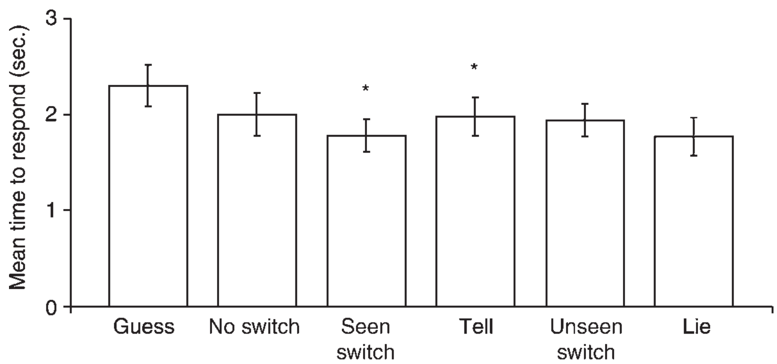

Figure 3. The mean latency to choose a location in each of the experimental conditions compared to the guess condition.

Table 4. Percentage of children showing different types of uncertainty across conditions and significance values associated with a comparison between the guess control condition and all the other conditions (McNemar test). Also shown is the overall score obtained after combining all uncertainty measures and the comparison between the guess control condition and all the other conditions (Wilcoxon test). *p* values are two-tailed.

	Uncertain words		Uncertain looks		Other non-verbal signs of uncertainty		Overall	
	%	*p* value	%	*p* value	%	*p* value	Score	*p* value
Guess	15.7		7.2		26.5		.49	
No switch	8.4	.24	3.6	.51	14.5	.052	.27	.013
Seen switch	3.6	.013	2.4	.22	4.8	.000	.11	.000
Tell	7.2	.092	1.2	.125	9.6	.004	.18	.000
Unseen switch	3.6	.013	1.2	.063	4.8	.000	.10	.000
Lie	6.0	.077	6.0	1.0	10.7	.026	.23	.015

non-verbal signs of uncertainty in the guess condition. Thus, they were not guessing in the experimental conditions – they responded quickly, without hesitation – even when failing.

Discussion

This study investigated children's strategies and patterns of responding during various versions of the traditional false belief test to determine whether children are truly passing the tests using an understanding of false belief. One aim of the study was to investigate whether children in the age range of $3\frac{1}{2}$–4 years – who show a chance level of performance as a group on false belief tasks – are guessing or deliberately choosing an answer. Replicating earlier findings (see Wellman *et al.*, 2001, for an overview), a little more than half the children of this age range failed the false belief task in our study. Results of various measures of uncertainty indicated that the children were not guessing. Children showed less uncertainty in the false belief conditions than in a condition in which they were forced to guess, and results for latency showed a similar trend. Further evidence that children were not guessing comes from the finding that passers and failers of the task did not differ in signs of uncertainty or latency.

These findings thus replicate the Carpenter *et al.* (2002) uncertainty findings with a traditional task. They also agree with the findings of Ruffman *et al.* (2001), if only the children in the same age group as in the current study are considered. That is, in the current study the age range of $3\frac{1}{2}$–4 years parallels closely the group of younger passers (mean age 3.6 years) and the group of older failers (mean age 4.1 years) in the Ruffman *et al.* study, and there was no difference in uncertainty between these two groups in that study.

Both passers and failers of the false belief task thus answered with a high degree of certainty – it is just that some of them were incorrect. This finding highlights the individual differences in performance in this age range.

An investigation of the patterns of individual children sheds some light on the differences between passers and failers. For example, although the unseen switch and lie conditions were of equal difficulty for children as a group, it was not always the case that the same children who passed the unseen switch condition also passed the lie condition, and performance on the two tests was not correlated with each other. Although this lack of relation between these two conditions may seem puzzling, other studies have also failed to find a positive relation between different false belief tests such as the location change and the unexpected contents task (e.g. Naito, 2003). Children with higher vocabulary and grammar scores were more likely to pass the unseen switch and/or lie conditions than children with lower vocabulary and grammar scores (although high language scores did not automatically lead to success). These results support other studies that have found a close relation between language development and the development of false belief understanding (Astington & Jenkins, 1999). Moreover, it should be noted that this connection shows up in the current study despite the use of rather unspecific measures of children's language development.[3] Interestingly, the same relation did not hold for the true variations of the test: children's language scores were not related to performance in any of the three true belief conditions.

Thus, while linguistic maturity was a predicting factor for passing the false belief tests, the true belief tests did not seem to rely on these linguistic factors (in this age range). Since the story scripts as well as the test questions were held equal in the different conditions in our study, these findings support the claim that the relation between language development and passing a false belief task is not simply caused by the language used in the task, but instead denotes a more fundamental and specific relation between language development and false belief understanding in preschool children. Another line of evidence comes from studies using non-verbal false belief tasks, which seem to be of equal difficulty (Call & Tomasello, 1999) or could be even more difficult (Plaut & Karmiloff-Smith, 1993) than verbal tasks. Compatible with the notion of a fundamental and specific relation between language development and false belief understanding is the finding of Surian and Leslie (1999) that it is only in a false belief (unseen switch) condition – and not in a true belief (seen switch) condition – that the insertion of a 'look first' question results

[3] *i.e. which do not include mental state terms or embedded sentences, which are commonly used to express belief situations, and which are hypothesized by some authors to play an important role in passing a standard false belief test (e.g. Astington, 2000; De Villiers, 2000; Lohmann & Tomasello, 2003)*

in a significantly higher success rate compared to the standard question. It could thus be that only children sensitive to fine distinctions in the linguistic input (i.e. children with advanced linguistic skills) profit from this information when solving a false belief task, while it makes no difference for solving a true belief condition, in which the protagonist is informed of the object's location.

With respect to theoretical assumptions regarding the factors (conceptual vs. processing) that influence children's developing understanding of false belief situations, we found evidence for both factors. In line with the developing concept of belief, children had more difficulties passing scenarios that involved a false belief than scenarios that involved true belief. Moreover, when comparing conditions of similar processing demands, we find negative or nonsignificant correlations between true and false belief conditions. This finding paired with the positive correlations found within true belief tasks speaks against the assumption that processing factors constitute a crucial element of the test. Nevertheless, processing demands influenced children's performance to some extent because when the scenario became more complex, the difference in performance between true and false conditions vanished. Moreover, within true belief conditions, adding a second possible location for the object (no switch vs. seen switch) led to more mistakes (performance dropped from 80% to 64%). Thus, language and conceptual skills, as well as processing demands influenced children's performance in the test.

Another important task demand may have been the pragmatics of the situation. It is interesting to note that although most children passed the no switch condition, surprisingly this condition elicited many signs of uncertainty compared to other belief conditions. Although still significantly lower than the guess condition with regard to uncertain looks or words, the no switch condition did not differ in latency to answer the test question or in other non-verbal signs of uncertainty from the guess condition. One explanation could be that some children were confused about how to interpret the intention behind the experimenter's test question when the solution left no real alternative options (because the object was put back in the same location). Other studies have found a similar pattern of inverted relations. For example (Russell, Hill, & Franco, 2001), found that older children (4-year-olds) found a false belief task slightly easier than a true belief task, while these tests were of equal difficulty for 3-year-olds. It appears that a true belief scenario can sometimes pose additional problems for more advanced children as they try to interpret the test question in a useful pragmatic way.

The findings from this study thus support hypotheses which involve the combined influence of processing factors as well as verbal development on performance on theory of mind tests for children of $3\frac{1}{2}$–4 years of age (e.g. Astington, 1990; Leslie & Thaiss, 1992). An interesting question for future research is the interaction of these two factors.

Another aim of this study was to determine whether children use a simple rule involving seeing instead of calculating a protagonist's belief when passing the standard false belief task. We attempted to tease apart the seeing-knowing confound inherent in the task by adding conditions in which the protagonist does not actually see the location of the object, but is later told about it (correctly or incorrectly). The results indicated that $3\frac{1}{2}$ to 4-year-old children can use both visual and verbal information when deciding where others will look for a toy. This suggests that when children pass false belief tests they do not use a simpler response strategy of: 'Protagonist will go where she last saw the toy' when passing the test. There is ample evidence that children as young as 2 (O'Neill, 1996) or 3 years of age understand something about the link between seeing and knowing in others. For example, 3- and 4-year-old children understand that a person who has looked into a box knows what is in the box but a person who has touched the box but not looked inside does not know (e.g. Pillow, 1989; Pratt & Bryant, 1990). This understanding allows children to create a theory of behaviour based on simple causal links between the mind and world. Incorrect answers in change-of-location tasks could conceivably due to using a desire link (see Wellman, 1990, for this approach), and correct answers due to using a perceptual link. But instead we showed that children were capable of updating others' visually-based knowledge with knowledge obtained through a different modality (i.e. verbal). Children can do this even when the information is false (the unseen switch and lie conditions were of equal difficulty). They thus appear to be using mental (belief/knowledge) information instead of perceptual information to solve the false belief task.

This study thus sheds some light on the performance of children who are at chance level as a group. Children in this age range are not randomly guessing, but deliberately choosing on the basis of belief, not on a simpler reasoning strategy. Influencing factors for test performance are the complexity of the test scenario and the child's verbal abilities.

Acknowledgements

We thank Michael Tomasello for his help in designing the study, and his comments on an earlier draft. We also thank Solvejg Kuehnert, Sonja Meyer, Susanne Mauritz, and Jana Jurkat for their help with coding and testing.

References

Astington, J. W. (1990). Narrative and the child's theory of mind. In B. Britton & A. Pellegrini (Eds.), *Narrative thought and narrative language*. Hillsdale, NJ: Erlbaum.

Astington, J. W. (2000). Language and metalanguage in children's understanding of mind. In J. W. Astington (Ed.), *Minds in the making: Essays in honor of David R. Olson*. Oxford, UK: Blackwell Publishers.

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

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition*, *21*, 37–46.

Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development*, *70*, 381–395.

Carpenter, M., Call, J., & Tomasello, M. (2002). A new false belief test for 36-month-olds. *British Journal of Developmental Psychology*, *20*, 393–420.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377–395.

DeVilliers, J. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind*. Hove, UK: Psychology Press.

Garnham, W. A., & Ruffman, T. (2001). Doesn't see, doesn't know: Is anticipatory looking really related to understanding of belief? *Developmental Science*, *4*, 94–100.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development*, *59*, 1366–1371.

Gopnik, A., Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New York: Cambridge University Press.

Grimm, H., & Schoeler, A. (1991). *Heidelberger Sprachentwicklungstest* (2nd ed.). Goettingen, Germany: Hogrefe Verlag.

Johnson, C. N., & Maratsos, M. (1977). Early comprehension of mental verbs: Think and know. *Child Development*, *48*, 1743–1747.

Johnson, E. (1997). Children's understanding of epistemic conduct in self-deception and other false belief stories. *Child Development*, *68*, 1117–1132.

Kaufman, A., & Kaufman, N. (1994). *Kaufman Assessment Battery for Children*. Frankfurt am Main, Germany: Swets and Zeitlinger test services.

Leslie, A. M, & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, *43*, 225–251.

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, *74*, 1130–1144.

Naito, M. (2003). The relationship between theory of mind and episodic memory: Evidence for the development of autonoetic consciousness. *Journal of Experimental Child Psychology*, *85*, 312–336.

O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, *67*, 659–677.

O'Neill, D. K., & Gopnik, A. (1991). Young children's ability to identify the sources of their beliefs. *Developmental Psychology*, *27*, 390–397.

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

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development*, *60*, 689–700.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*, 125–137.

Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology*, *47*, 116–129.

Pillow, B. H. (1993). Preschool children's understanding of the relationship between modality of perceptual access and knowledge of perceptual properties. *British Journal of Developmental Psychology*, *11*, 371–389.

Plaut, D. C., & Karmiloff-Smith, A. (1993). Representational development and theory-of-mind computations. *Behavioral and Brain Sciences*, *16*, 70–71.

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, *61*, 973–982.

Robinson, E. J., & Mitchell, P. (1992). Children's interpretation of messages from a speaker with a false belief. *Child Development*, *63*, 639–652.

Roth, D., & Leslie, A. M. (1998). Solving belief problems: Toward a task analysis. *Cognition*, *66*, 1–31.

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology*, *80*, 201–224.

Russell, J., Hill, E. L., & Franco, F. (2001). The role of belief veracity in understanding intentions-in-action. Preschool children's performance on the transparent intentions task. *Cognitive Development*, *16*, 775–792.

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition*, *38*, 1–12.

Surian, L., & Leslie, A. M. (1999). Competence and performance in false belief understanding: A comparison of autistic and normal 3-year-old children. *British Journal of Developmental Psychology*, *17*, 141–155.

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

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

Wimmer, H., Hogrefe, G., & Perner, J. (1988). Children's understanding of informational access as source of knowledge. *Child Development*, *59*, 386–396.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and 'false' photographs. *Cognition*, *35*, 41–68.
Zaitchik, D. (1991). Is only seeing really believing?: Sources of the true belief in the false belief task. *Cognitive Development*, *6*, 91–103.

Received 24 July 2003; revised version received 12 July 2004

Appendix

A script of each of the conditions in one of the storylines

Note. Text in italics was presented to all children; text below the condition name was specific to that experimental condition.

Protagonist 1: Hi, I am Lisa. I am here in my room. Look, here is a bed with a drawer (location A). And look here is also a box (location B). Can you see them? Look, here I have a necklace (object). Mmm, how beautiful. Now I want to put the necklace away. I will put the necklace in the drawer (location A).

No switch (True belief)

Protagonist 1: Now I will go play. (leaves)
Protagonist 2 appears: Hi, I am Lisa's sister, Anna. Lisa is not here anymore. Ah, here is a drawer (location A). And here is a box (location B). Let's see, what is in here (location where necklace is). Oh, there is a necklace! I'll get it out. Oh, how beautiful! Now I'll put the necklace back into the drawer (old location). Now, I'll go out to play. (leaves)

Seen switch (True belief)

Protagonist 1: Then I will sit down a bit. (sits on bed)
Protagonist 2 appears: Oh, hi, Lisa, you are also here.
Protagonist 1: Hi Anna!
Protagonist 2: Ah, here is a drawer (location A). And here is a box (location B). Let's see, what is in here (location where necklace is). Oh, there is a necklace! I'll get it out. Oh, how beautiful. Now I won't put the necklace back into the drawer, instead I will put it into the box. Now, I'll go out to play. Bye bye Lisa. (leaves)
Protagonist 1: Bye bye.

Tell (True belief)

Protagonist 1: Now I will go play. (leaves)
Protagonist 2 appears: Hi, I am Lisa's sister, Anna. Lisa is not here anymore. Ah, here is a drawer (location A). And here is a box (location B). Let's see, what is in here (location where necklace is).Oh, there is a necklace! I'll get it out. Oh, how beautiful. Now I won't put the necklace back into the drawer, instead I will put it into the box. Now, I will go and look for Lisa. (Lisa appears). Hi, Lisa! I moved the necklace to a new place. I put the necklace into the box. Bye bye. (leaves)
Protagonist 1: Alright! Bye bye!

Unseen switch (False belief)

Protagonist 1: Now I will go play. (leaves)
Protagonist 2: Hi, I am Lisa's sister, Anna. Lisa is not here anymore. Ah, here is a drawer (location A). And here is a box (location B). Let's see, what is in here (location where necklace is). Oh, there is a necklace! I'll get it out. Oh, how beautiful. Now I won't put the necklace back into the drawer, instead I will put it into the box. Now, I'll go out to play. (leaves)

Lie (False belief)

Protagonist 1: Now I will go play. (leaves)
Protagonist 2: Hi, I am Lisa's sister, Anna. Lisa is not here anymore. Ah, here is a drawer (location A). And here is a box (location B). Let's see, what is in here (location where necklace is). Oh, there is a necklace! I'll get it out. Oh, how beautiful. Now I'll put the necklace back into the drawer (location where it was). Now, I will go and look for Lisa. (Lisa appears). Hi, Lisa! I moved the necklace to a new place. I put the necklace into the box! Bye bye. (leaves)
Protagonist 1: Alright! Bye bye.

Guess control

Protagonist 2: Hi, I am Anna. I am the sister of Lisa. I am here in her room. Look, here is a bed with a drawer (location A). And look, here is also a box (location B). Can you see them? Look, here I have a necklace. It belongs to my sister Lisa. She wants to get it later. Now I want to put the necklace away.
-------Bus drives through the scene, covering it for a moment------
I have put the necklace away. I put the necklace either in the drawer (location A) or in the box (location B).
Now, I'll go out to play. (leaves)
Protagonist 1: Hello, I am Lisa.
Protagonist 1 (reappears): Hello! I want to have my necklace now!

Control questions: Where did Lisa (Anna in the Guess condition) put the necklace in the beginning? Where is the necklace now?

Test question: Where will Lisa look for the necklace first?