Theory of Mind and Self-Control:
More than a Common Problem of Inhibition

Josef Perner, Birgit Lang, and Daniela Kloo

This study tested the theory that advances on theory-of-mind tasks and on executive function tasks show a strong correlation because the typically used theory-of-mind tasks pose the same executive demands. In Experiment 1 with fifty-six 3- to 6-year-old children, performance on the dimensional change card-sorting task as an executive function task was correlated with performance on the usual false-belief prediction task, *r* = .65, and the false-belief explanation task, *r* = .65, as measures of theory-of-mind development. Because the explanation version of the false-belief test is supposed to be free of the alleged executive demands inherent in the prediction version, the equally strong correlation with the executive function task suggests that this correlation cannot be due to common executive demands. In Experiment 2, the basic finding of Experiment 1 was replicated on another sample of 73 children, ages 3 to 5.5 years. The need for new theories to explain the developmental link between theory of mind and executive function development is discussed, and some existing candidates are evaluated.

INTRODUCTION

It was first noted by Russell, Mauthner, Sharpe, and Tidswell (1991) and Hughes and Russell (1993) that the ability to understand false belief, a critical stage in the acquisition of a theory of mind, emerges with improvements on executive tasks, which are measures of children's growing self-control. This developmental relation has now been confirmed in several studies using different tests to assess children's understanding of mental perspective (involving false belief, appearance–reality distinction, and deception) and a host of different executive function tests: wrong pointing, switching to new instrumental methods or new sorting dimensions, suppressing imitative tendencies, and so forth (Carlson & Moses, 2001; Davis & Pratt, 1995; Frye, Zelazo, & Palfai, 1995; Gordon & Olson, 1998; Hughes, 1998a, 1998b; Hughes & Russell, 1993; Hughes, Dunn, & White, 1998; Keenan, 1999; Russell et al., 1991). A brief meta-analysis of these studies by Perner and Lang (2000) showed an effect size for the reported correlations of 1.08, which is generally considered a strong effect (Glass, McGaw, & Smith, 1981).

The first and most plausible explanation for this developmental relation (Hughes & Russell, 1993; Russell et al., 1991) was that the observed correlations were due to executive components in the false-belief and related tasks. For instance, in the traditional belief task (Wimmer & Perner, 1983), a story character does not witness the unexpected transfer of an object from its original location (A) to a new location (B). When asked where that person will look for the object, children have to exert internal executive control over the habitual schemas evoked by current reality (salient reality) in favor of the protagonist's belief world (Harris, 1993; Russell et al., 1991). Moreover, children have to disengage their attention—which is naturally captured by the location of the desired object—from that location and focus it on the location of where the protagonist thinks the object is (Hughes & Russell, 1993); and they have to suppress the natural but wrong prediction that a person typically looks for an object where it is (Carlson, Moses, & Hix, 1998). Unless these natural tendencies can be kept in check then, despite the capacity for a deeper understanding, children wrongly answer that the protagonist will look in the object's actual location.

Similarly, in the appearance–reality task (Flavell, Flavell, & Green, 1983), which is another popular measure of theory-of-mind development, children discover by touching that an object that deceptively looks like a rock is actually a piece of sponge. When asked what the object "looks like," children have to inhibit salient reality and the natural tendency to blurt out what it actually is.

There is also some direct empirical support for this theory in connection with children's ability to deceive as a measure of their theory-of-mind development (Carlson et al., 1998). When asked where an object is, children find it easier to deceive by turning a pointer to an empty location (because this may help them to reflect and because there is no automatic response tendency for operating pointers) than by either answering with a wrong location or by pointing to it

with their finger (response modes for which there are automatic tendencies).

It should be noted that an explanation at the level of an automatic response tendency is much more plausible in the case of deceptive pointing than in the case of the false-belief or the appearance–reality task. When asked where something is located, there is, indeed, a very strong tendency to automatically point to or answer with the actual location. This tendency needs to be inhibited if one wants to deceive. This is different in the false belief story. Children are asked where a mistaken person will look for an object. If children understand that this person will look in the wrong place, then it is difficult to accept the suggestion that they tend to answer naturally and automatically with a different location. In this instance, the additional assumptions voiced first by Russell et al. (1991) come into play; that is, children are so preoccupied with the object's real location that they tend to answer with that location, or the real state of affairs is cognitively so salient that it dominates the children's mental computations.

Several other authors also have emphasized the need to inhibit salient reality in favor of beliefs about reality. The theory that salient reality masks young children's true understanding of belief (Mitchell, 1996) has received some support. Children answer more often correctly in terms of the story protagonist's belief when reality is invisible (the target object is hidden) than when it is perceptually salient (Zaitchik, 1991). Furthermore, performance is better when the belief contents are made visible (in a picture or symbol; Mitchell & Lacohée, 1991) than when simply inferred by children. Leslie and Polizzi (1998) and Fodor (1992) saw the problem slightly differently—as one of inhibiting a default attribution response, namely the default that beliefs are true.

One piece of evidence that is often referred to by proponents of these different versions of the inhibition theory of performance on theory-of-mind tasks is the claim made by Bartsch and Wellman (1989) that children are able to give correct explanations of erroneous actions due to a false belief well before they make correct action predictions in the standard belief task. The saliency hypothesis by Russell et al. (1991) predicts this because incorrect explanations are not due to (uninhibited) reference to the desired object's known location. Mitchell (1996) points out that the protagonist's wrong search in the explanation paradigm provides the needed reality counterpart to the protagonist's false belief to help children overcome the misleading pull of the object's real location; and Fodor (1992) argued that the explanation paradigm should be easier because it makes clear that the default assumption of correct behavior is not warranted.

The interpretation of Bartsch and Wellman's (1989, Experiment 2) finding has, however, been questioned on several fronts. In their experiment, children were shown an empty Band-Aid™ container and told that puppet Bill has a cut and wants a Band-Aid. The children then watched Bill look inside the empty Band-Aid container and were asked why he was looking in there. Some children spontaneously gave sensible explanations. As Astington (1996) has pointed out, however, there were only about as many of these correctly explaining children as there were children who could make correct predictions in the standard prediction task. Research by Wimmer and Weichbold (1994) and Wimmer and Mayringer (1998) confirmed that children find open-ended explanation questions as difficult as prediction of an erroneous action.

Bartsch and Wellman's (1989) case rested exclusively on additional "correct" answers in response to a helpful prompt—for example, when children failed to answer the original question, Bartsch and Wellman then asked, "what does Bill think?"—after which a substantial number of additional children gave relevant-sounding answers referring to "Band-Aids." However, these prompted answers could, plausibly, have been false positives because some children at this age might answer "think" questions in terms of what someone wants, as children do tend to take "think" information as an indication of what a person wants (Wellman & Bartsch, 1988, Experiment 1). These answers could also have been false positives because the empty Band-Aid box did not suggest any alternative but "Band-Aids" when children were pressed for a plausible answer (Wimmer & Mayringer, 1998). When such an alternative was provided (e.g., rocks in the Band-Aid box) then there was no real advantage of the explanation paradigm over the prediction paradigm (Moses & Flavell, 1990, Experiment 1). Moreover, in their second experiment, Moses and Flavell (1990) found that many of the younger children who gave seemingly correct answers to the question about what the protagonist (as he was about to look for Band-Aids in the Band-Aids box) was thinking, kept giving the same answer after the protagonist had discovered that it was filled with rocks. This answer, now evidently wrong, suggests that these children may have assimilated "what does he think is in the box" to "what is he pretending is in the box" (Perner, Baker, & Hutton, 1994).

To circumvent the methodological chasm of asking open-ended explanation questions to avoid false positives and trying to give children a chance to show their true competence with helpful hints, Robinson and Mitchell (1995) devised a new task variant. Two identical-looking twins place their ball into one (A) of

two boxes. One of the twins then leaves the room, whereupon the other twin takes the ball out of Box A, plays with it some more, and then places it into the other Box (B) before leaving the room. Later both twins return with their mother who asks them about the ball. One of them goes to Box A, the other to Box B. Children are asked why the one twin goes to Box A: "Was it because he had left early or because he had been playing with the ball some more?" Although it is possible that false positives occur because some children simply associate absence from the ball with going to the empty container, it is unlikely that this would affect many children. Thus, it is remarkable that a large difference was found between correct predictions in a prediction task and correct choice of explanation in the twins task (45% versus 87% in Experiment 5 and 15% versus 67% in Experiment 6).

This difference, however, can be explained almost entirely by a difference in baseline (Perner, 1995). The plausible assumption is made that a failure to understand belief results in pure guesses on the forced choice of the explanation task (i.e., about 50% correct) but results in predominantly incorrect answers on the prediction task (i.e., near 0% correct) because children tend to apply the incorrect theory that the person will look where the object really is. With example data from Robinson and Mitchell's (1995) Experiment 6, it is possible to reason as follows: 15% of the children understood belief because that many made correct predictions; therefore, a corresponding 15% of children gave correct explanations because they understood belief. Because by assumption, the remaining 85% simply guessed on the explanation choices (42.5% correct by chance), a total of 57.5% correct explanations can be expected. This means that the observed frequency of 67% correct explanation choices reflects a mere 24.5% (i.e., 67% − 42.5%) that were due to true understanding, which is a mere 9.5% more than the 15% correct predictions that were observed.

One objective of the present study was to check the validity of this post hoc analysis, by contrasting children's performance on several identical-looking twin explanation tasks (in different story guises) to gauge the incidence of guessing. If the analysis is correct, then children who fail the prediction tasks will show a guessing distribution, whereas those who pass the prediction tasks will give mostly correct explanation choices. This result would speak against the claim that the prediction task is more difficult than the explanation task, because prediction involves more of an executive, inhibitory problem than does explanation.

There could, however, be more genuinely correct explanation choices than predictions. If this is true,

then of the children who fail the prediction task, some will make systematically correct choices (because they understand), whereas others will have systematically wrong choices on the prediction task (e.g., because they prefer to opt for the choice linking the twin to the desired object: he is looking for the ball in this [empty] box, "because he stayed in the room and played with the ball"). Such a finding would support the contention that explanation is easier than prediction and that this difference may be due to stronger executive demands in the prediction task.

A second, related objective was to test directly whether performance on an executive test correlates only with the prediction version of the false-belief task (due to the need to inhibit salient reality) or also with the explanation version (which does not have this particular executive demand). For this purpose, the present study used the dimensional card-sorting task (Frye et al., 1995), which is an adaptation for use on children of the Wisconsin card-sorting task that is traditionally used to diagnose executive problems in frontal lobe patients. If the reported correlations between theory-of-mind development (typically tested with the false-belief prediction task) are due to the need to inhibit salient reality, then there should be no, or a clearly reduced, correlation with the explanation task. If, however, there is a deeper developmental link between executive tasks and understanding belief, then the correlations should be similar. Hughes (1998a) reported that executive tasks correlated as strongly with an explanation version as with a prediction version. It is still possible, however, that even children's free answers to the explanation question are subject to the executive demand of having to inhibit the tendency to mention the desired object or to refer to its place (a not-infrequent wrong answer, Wimmer & Mayringer, 1998). The use of forced choices in the twin task allowed for the ability to test the presence of such an answer tendency.

EXPERIMENT 1

Method

Participants

Fifty-six children (33 girls, 23 boys) between the ages of 3,1 and 6,2 ($M = 4,5$, $SD = 10.3$ months), from a large kindergarten in a middle- and working-class area of Salzburg, Austria, volunteered for this study. All children within this age range who were available in the kindergarten class at the time of this study participated. There were 22 children between 3,1 and 3,11, 18 children between 4,1 and 4,11, and 16 children between 5,0 and 6,2.

Design

Each child was tested in two sessions about 1 week apart. Testing per session lasted between 10 and 20 min. Each session consisted of four tasks, which were presented in eight different sequences with 7 children per sequence. In Positions 2 and 4 in the first session and Positions 1 and 3 in the second session the four twin scenarios of the false-belief tasks were used: three as explanation tasks, and one as a prediction task. Which twin scenario was given as the prediction task was systematically varied through the four positions. The other four positions contained the following tasks: a traditional false-belief scenario (prediction version), two-dimensional change card-sorting tests (DCCS), and the Kaufman Assessment Battery for Children (K-ABC; Melchers & Preuß, 1991), a verbal intelligence test. For four sequences, the card-sorting task was in Position 1; for two sequences, the intelligence test was in Position 3; and for the other two sequences, the traditional false-belief test was in Position 3. For the other four sequences, the false-belief or intelligence test was in Position 1 and the card-sorting task was in Position 3. These sequences were repeated for the second test session (in Positions 2 and 4) with the exception that when the intelligence test was used in the first session the belief test was given in the second session and vice versa.

In sum, each child was given the K-ABC verbal intelligence test, two card-sorting tasks (DCCS), three false-belief explanation tasks, and two false-belief prediction tasks of which one was the more traditional task and the other was couched in one of the four twin scenarios.

Procedure and Materials

Card Sorting

Two sets of cards (10 cm × 7 cm) were used. Each of these card sets consisted of two target cards (a big yellow square and a little red square; a red cat and a black snake) that were each affixed to a (26 × 14 × 12 cm) box into which the test cards had to be posted through a slit. There were 24 test cards: 6 with big red cars, 6 with little yellow cars, 6 with black cats, and 6 with red snakes. The card-sorting task in each session involved a preswitch phase and a postswitch phase.

Preswitch phase. The experimenter explained the two dimensions (color and size) of the target cards. For instance, using the card set with big red and little yellow cars, the experimenter explained, "Now we are playing a game, a COLOR GAME. In this game, all the red cards go here, but the yellow ones go in the other box." The children and the experimenter sorted two cards together (one red and one yellow) and then the children were required to sort five cards on their own. On each of these five preswitch trials the experimenter repeated the preswitch rules to the children. She randomly selected a test card and labeled the card with the relevant dimension (e.g., "Here is a red one"). Then the children were asked to place the card in one of the two boxes ("Where does this card go in the color game?") and they were told whether they had sorted the card correctly.

Postswitch phase. After the fifth preswitch trial children were told: "Okay, now we are going to play a new game, the SIZE GAME. The size game is different. This time, all big cars go here, but all little cars go there." Again, the children had to sort five cards, but according to the new size rules and without feedback as to whether each card was placed correctly. Every time a card was placed incorrectly, however, the rule was repeated before the next trial. The card-sorting task in the second session was the same except that a different card set was used that varied according to color (black versus red) and shape (cat versus snake). In both sessions, the first dimension was always color and the second was either size (first session) or shape (second session).

False-Belief Task

Explanation tasks (twin scenarios). Following the procedure of Robinson and Mitchell (1995) four different stories with identical-looking twins were enacted. In the ball story the twins are playing ball in a room and then place the ball into one of two boxes. One twin then leaves the room to have a glass of milk in the kitchen. In his absence the other twin plays with the ball for awhile and then places the ball into the other box and leaves the room as well. Then both twins return with their mother who asks them where their ball is. One twin goes to the box where the ball actually is, the other twin goes to the other box. Children are then asked about the mistaken twin: "Why does this one look in this box?" Children's answers to this open question are noted, but regardless of their answer they are given a forced choice: "Is it because he had stayed in the room and played with the ball or is it because he had left the room in order to drink some milk in the kitchen?" The verbatim narrative is given in the Appendix.

The other scenarios are structurally identical but the story material and the protagonists are different. For example, the chocolate story is the same except that instead of a ball a chocolate bar is being transferred from one location to the other. The bird and the rabbit stories are different in that the bird and the rab-

bit move on their own accord. For instance, in the bird story the children are playing in the garden and they see a bird. One of the twins decides to fetch some bird feed. The other twin is watching the bird and witnesses the bird leaving its nest and flying behind a shrub.

Prediction task. A traditional false-belief task (Wimmer & Perner, 1983) was administered with a dog as protagonist. In the dog's absence his bone was unexpectedly transferred to a new location. Children's understanding of the dog's false belief was assessed by asking them to predict where the dog would look for his bone. Their memory of critical story events was assessed by two control questions about (1) where the dog had put his bone in the beginning and (2) where the bone was now. The story was enacted in a three-dimensional model (42 cm × 44 cm × 78 cm) with two kennels that had different colored roofs (green and yellow). A dog made of fabric and a puppet representing the dog's master (who transferred the bone from one kennel to another) were used as the main story characters; the bone was 2.5 cm × 1 cm and made of modeling clay.

For the second prediction task, one of the twin scenarios described above was adapted. The ball story, for instance, was changed in the following way: The twins were not identically dressed. The story was the same except that when the mother asked the twins where the ball was the experimenter pointed to the twin that had left the scene first and said, "Look! This is Peter who had left the room in order to have a glass of milk in the kitchen. Where will Peter go to look for the ball?"

Verbal Intelligence Test (K-ABC)

To ensure that any correlation found between the tasks was not due just to increases in verbal intelligence, the vocabulary subtest of the K-ABC (Melchers & Preuß, 1991) was administered. The test consists of 24 pictures (e.g., dog, TV set, scissors, and thermos) that children have to identify.

Results and Discussion

Theory of Mind

False-Belief Prediction

Only 9 children gave a wrong answer on one of the two control questions. Performance on the control questions was partialed out, and it was found that the interpretation of the main results was not affected. Answers to the test questions were somewhat better on the traditional version (66% correct) than the twin scenario (51.7%): McNemar's $\chi^2(1, N = 12) = 5.33, p < .05$. Nevertheless, the two tests correlated fairly strongly, $r = .59$, yielding moderate (Shrout, 1998) retest reliability, $\kappa = .57$.

False-Belief Explanation

Children's answers to the open question in the three twin scenario tasks were classified according to the following categories: (1) mental state, 44 answers (e.g., "He thought it was in there," "he doesn't know it's in the other cupboard," or "he didn't see it being moved."); (2) relevant story facts, 32 answers (e.g., "He had been away," or "it was in here earlier."); (3) desire, 18 answers (e.g., "because he wants the ball."); (4) wrong location, 31 answers, (e.g., "because the ball isn't in here," or "because the ball is over there."); (5) irrelevant facts, 17 answers (e.g., "I want to play something else now," "Where is the bird?", or "Why?"); and (6) no or "don't know," 26 answers. In the case of multiple answers, the one that fit the "best" category was used; for example, if a child said "Don't know— 'cause he wants the ball," this was classified as Category 3 and not Category 6.

For further analysis, answers in Categories 1 and 2 were classified as "correct answers" indicating an understanding of belief. Answers in the remaining categories were classified as "incorrect answers," indicating failure to show such understanding (Wimmer & Mayringer, 1998). Twenty-five children gave no correct answer, 11 gave one or two correct answers, and 20 gave correct answers on all three tasks, which was a strongly bimodal distribution (comparing zero and three correct answers to one and two correct answers), $\chi^2(1, N = 56) = 20.64, p < .001$. Correspondingly, the three pairwise correlations between the three administrations of the task, $.72 \leq r \leq .78$, and the retest reliabilities for the three pairings of tasks, $.71 \leq \kappa \leq .78$, were very high.

Table 1 shows how answers to the open question related to the subsequent forced-choice answers. Reassuringly, practically all the correct answers to the open question were followed by correct answers to the forced choice. On only two occasions did children fail to give any answer and on only six occasions did they give the wrong answer. In contrast, when an uninformative or wrong answer was given to the open question, children frequently (12 times) did not give any answer or they gave a wrong answer to the forced-choice question (28 times). They still gave a correct answer more frequently (52 times) than a wrong answer, however, thus supporting the original suspicion that reliance on answers to open ques-

Table 1   Number of Stories in Which Children Gave Particular Combinations of Answers to Open Questions and Forced-Choice Questions

	Answer to Forced-Choice Question		
Answer to Open Question	Correct	Incorrect	No Answer
Correct			
Mental state	38	4	2
Relevant story facts	30	2	0
Incorrect			
Desire	12	6	0
Wrong location	19	10	2
Irrelevant facts	6	8	3
No/don't know	15	4	7

Table 2   Number of Children Who Gave Correct Answers to Forced-Choice Explanation Questions and Prediction Questions in Experiment 1

No. of Correct Predictions	No. of Forced-Choice Explanations and No. Correct Thereof										No. of Correct Free Explanations				
	3				2			1		0					
	0	1	2	3	0	1	2	0	1	0	0	1	2	3	Total
0	0	3	5	2	1	2	1	2	0	1	13	3	0	1	17
1	1	4	1	4	0	0	2	0	1	1	7	0	3	2	12
2	0	1	3	22	0	0	1	0	0	0	5	3	2	17	27

tions may underestimate children's true explanation competence.

Interestingly, the fact that more than half (49) of all wrong answers (92) fell into Category 3 (desire) and Category 4 (wrong location) raised the possibility that children's free answers to the open question were hampered by the executive problem of having to disengage attention from the desired object, and confirmed the need for the helpful clue given in the forced-choice question that might have helped children to focus on the relevant options.

The pairwise correlations between forced-choice answers were considerably lower, $.27 \leq r \leq .45$, than were the correlations for the free answers to open explanation questions and for predictions. This was to be expected, because it is likely that on this measure the younger children, who did not understand false belief, simply guessed one of the two alternatives. That such guessing did indeed occur was also confirmed in the following analysis.

Relating Prediction and Explanation

Of main concern for the present study's purposes was how children's forced choices on the three explanation tasks related to their answers on the two prediction tasks. This analysis was complicated by the fact that not every child answered all three forced-choice questions. Table 2, therefore, details response frequencies according to how many of the forced-choice questions were answered.

The results given in the last column of Table 2 confirm the (well-known) fact that on prediction tasks children tend to give systematically incorrect answers. If the 17 children who gave incorrect answers on both tasks (i.e., zero correct, first row) had been purely guessing then there should not have been 17 in

that category but rather—according to the binomial distribution—only half of the 12 children who guessed once correctly (second row); that is, only about 6. Moreover, in most cases giving only one correct answer was not the product of guessing, because 10 of the 12 children who gave one correct answer made an error on the twin scenario prediction task and only 2 children made an error on the traditional task. That means that very few, if any, of these children got one answer right due to guessing. Rather they gave only one correct answer because they understood the easier version and gave consistently incorrect answers on the more difficult version.

The picture looks similar for the forced-choice explanation answers when the 27 children who succeeded on both prediction tasks are considered (third row in Table 2): All children but 1 answered at least two of the three forced-choice explanations correctly—a significant difference from what would be expected if children had simply been guessing, Sign test $\chi^2(1, N = 27) = 25.0$, $p < .001$. The children who showed some understanding of belief on one of the prediction tasks (second row) tended to do some guesswork, with a bias toward answering more questions correctly than incorrectly—but not significantly so, Sign test $\chi^2(1, N = 13) = .69$, $p > .30$. With regard to the children who showed no understanding of belief on the prediction tasks (first row), there was no sign of any consistency in explanation, Sign test $\chi^2(1, N = 14) = .14$, $p > .70$. They seemed to be largely guessing, with a possible slight tendency toward giving more correct than incorrect answers (i.e., 8 children gave more correct than incorrect answers, whereas only 6 gave more incorrect than correct answers). In sum, considering that 1 child did not give an answer to any of the three forced-choice questions, evidence to support the fact that children who failed prediction questions were able to show understanding of

belief with forced-choice explanation questions was minimal.

This result demonstrated two important points. It showed that when guessing was controlled for on forced-choice explanation questions in the twin paradigm, by and large, children who gave above-guessing level correct answers were also able to make correct predictions. Hence, there was little evidence that children's understanding of belief was masked by executive demands in the prediction task of having to inhibit salient reality of where the desired object really was. It also showed that in the forced-choice answers there was no strong tendency to give the wrong answer, for example, that children chose the option that mentioned the desired object (the boy who kept playing with the ball). Rather, the children who failed to understand belief as indexed by the prediction task showed a near-random choice of options (slightly biased toward the correct option). Performance on the forced-choice question was, therefore, a suitable test case for the issue at hand, namely, whether the correlation between false-belief tests and executive inhibition tests is based purely on executive demands in the false-belief test.

Table 2 also provides the contingency between prediction tasks and free answers to the open explanation questions. Overall there was good correspondence. Most children who made two correct predictions gave sensible answers to the explanation question, whereas only a few children who failed both prediction tasks did so. If the panel is dichotomized into children who completely failed on all tasks and those who gave at least one correct answer then 12 children (5 + 7) made at least one correct prediction without a single correct explanation, whereas only 4 children (3 + 1) gave at least one correct explanation while failing both prediction tasks. This difference is significant, McNe-

mar's $\chi^2(1, N = 16) = 4.0$, $p < .05$, and confirms the impression that prediction is a slightly more sensitive measure for detecting children's understanding of belief than are free answers to open explanation questions.

Card Sorting (DCCS)

The variable of interest in the card-sorting task was the number of correct responses after the switch to the new sorting rule. The majority of children sorted either five times correctly or five times incorrectly; that is, they continued according to the old rule. For the switch from color to form, 11 children had none correct, 8 had one to four correct, and 37 had all five correct; for the switch from color to size, 18 children had none correct, 7 had one to four correct, and 31 had five correct. Also, the switch from color to form was easier than the switch from color to size. There were 24 children who had more postswitch errors on color to size than on color to form but only 5 children who demonstrated the reverse, Sign test $\chi^2(1, N = 29) = 12.45$, $p < .001$. The correlation based on the continuous measure of items sorted correctly was substantial, $r = .78$.

Relating False-Belief Understanding to Card Sorting

Table 3 shows the basic statistics for each variable of interest and their intercorrelations. For each measure, the original continuous variable was used rather than a dichotomized version. The correlation coefficients (above the main diagonal) replicated the finding by Frye et al. (1995) that the false-belief prediction test strongly correlates with card sorting. The finding of central interest was that the correlation for the forced-choice explanation test had practically the same

Table 3   Basic Statistics and Correlations

| Variables | Mean % (SD) | K-ABC | False Belief | | | | | Card Sorting |
			Control Question	Prediction	Explanation Free	Explanation Forced	
Age (months)		.66**	.29*	.70**	.56**	.57**	.52**
K-ABC	60.0 (16.7)		.38**	.57**	.64**	.58**	.54**
FB control question	91.9 (18.5)	—		.43**	.40**	.21	.34*
FB prediction	58.9 (43.8)	—	—		.58**	.62**	.65**
FB explanation free	45.1 (45.5)	—	—	[.22]		.58**	.57**
FB explanation forced	71.2 (33.9)	—	—	[.37**]	[.29*]		.65**
Card sorting	68.4 (41.3)	—	—	[.40**]	[.29*]	[.47**]	

Note: Correlations in brackets below the main diagonal are after age, Kaufman Assessment Battery for Children [K-ABC], and control questions were partialed out. FB = false belief.
*$p < .05$; **$p < .01$.

highly significant correlation with the card-sorting task. In addition, children's answers to the open explanation question correlated highly with card-sorting. Moreover, these correlations remained highly significant even after age, verbal intelligence, and performance on the control questions had been partialed out (correlations in brackets below the main diagonal).

Because the error pattern on the forced-choice explanation task clearly established that no strong, wrong response tendency existed, the strong correlation with the card-sorting task established that there was a correlation between understanding false-belief and executive tasks that could not be explained by demands of executive inhibition in the false-belief task.

Another question of interest was whether the correlations between the different measures of understanding false belief and card sorting are based on the same or on different sources of variance. This issue was investigated by partialing out each false-belief variable in turn, with square of the partial correlation coefficient indicating how much of the card-sorting variance is explained by the remaining variables. For instance, the correlation between forced-choice explanations and card sorting of .65 indicates that 42.9% of the card-sorting variance is explained by forced-choice answers. By partialing out performance on the prediction tasks, this correlation reduces to a partial correlation of .42 corresponding to 17.7%, $p \leq .01$, of the card-sorting variance explained by forced-choice answers independent of the variance explained by the prediction task. The difference of 25.2% in percentage variance explained, $p \leq .01$, is the percentage of card-sorting variance attributable to a source common to the prediction and the forced-choice explanation tasks. Similarly, the corresponding computations for card-sorting variance explained by the prediction task independent of forced-choice explanations yields 16.0%.

These percentages could mean that 16% of the card-sorting variance could be explained by the prediction task due to a common inhibition demand. However, 25% of the variance is explained by a feature that is shared by prediction and explanation tasks, and it is unlikely that this could be based on the inhibition of a prevalent response or salient reality. The corresponding value for prediction task and answers to open explanation questions is 22.7%.

EXPERIMENT 2

One objective of this experiment was to investigate whether the central finding of Experiment 1, of a correlation between the explanation version of the false-belief task and card sorting, could be replicated using the traditional false-belief stories (instead of the somewhat more complicated twin stories) for explanation and prediction versions.

The second objective was to check whether the link between these tasks was due to the development of inhibitory abilities by introducing a fairly pure measure of inhibition: the false alarm rate on the go-nogo task (Luria & Tizard, 1961). This task requires children to press a button in response to one stimulus (e.g., a yellow light) and inhibit this response when a different stimulus appears (e.g., red light). Various versions of this task have confirmed a developmental change (Livesey & Morgan, 1991; Miller, Shelton & Flavell, 1970) in the age bracket of 3 to 6 years, which was of concern in the present study. This task, however, has not been used in connection with children's developing theory of mind. Theory-of-mind development has been linked to other inhibitory tasks, notably by Carlson and Moses (2001), but their tasks have a more complicated logical structure akin to that of the card-sorting task in which the response (consequent: c_i) to a stimulus (antecedent: a_i) needs to be made dependent on some background setting (Frye et al., 1995): <Setting 1: $a_1 \rightarrow c_1$, $a_2 \rightarrow c_2$; Setting 2: $a_1 \rightarrow c_2$, $a_2 \rightarrow c_1$>. The go-nogo task has the simpler rule structure of <$a_1 \rightarrow$press, $a_2 \rightarrow$do not press>, and failure to inhibit the press on stimulus a_2 appears to be a pure problem of inhibition uncontaminated with failure of representing background setting conditions or implementing rule switches.

Experiment 2 used three versions of the go-nogo task, each comprising 25 trials. Two of these versions were with 50% go and 50% nogo trials, one with an interstimulus interval (ISI) of 1 s and the other with an ISI of .5 s, to gauge children's problems with the speed of presentation. The third task used 75% go and 25% nogo trials with an ISI of .5 s, to increase children's inhibition problems and thereby their false alarms at the cost of correct inhibition (or correct rejections in signal-detection theory terms). We expected that children's correct inhibitions would increase with age. This experiment also provided the opportunity to test whether performance on the false-belief tasks (explanation and prediction) and card-sorting task would correlate with children's correct inhibition rate and whether a common correlation with the correct inhibition rate would account for the correlation between false-belief understanding and card sorting.

Method

Participants

All available children ($N = 82$; 50 girls, 32 boys) within the appropriate age range from seven small

kindergartens in Salzburg, Austria, and two kindergartens in villages near Salzburg were recruited. The children were from predominantly middle-class and upper working-class backgrounds. Nine children were unable to complete both sessions and were therefore dropped from the study. The final sample of 73 children consisted of children between the ages of 2,9 and 5,8 (M = 3,11, SD = 6.3 months).

Design

Each child was tested individually in two sessions about 1 week apart. The first session lasted about 40 min, and the second session lasted about 25 min. In each session, the following tasks were administered in a fixed order: a false-belief task, a "synonyms task" (Doherty & Perner, 1998; used for exploratory purposes in a different research project), card-sorting task (DCCS), one go-nogo task, and a second false-belief task (in each session there was one prediction and one explanation version in counterbalanced order). Moreover, in the first session, a German translation of the British Picture Vocabulary Scale (BPVS-II; Dunn, Dunn, Whetton, & Pintillie, 1982) was administered as a measure of verbal intelligence. It was given between the go-nogo task and the second false-belief task, whereas in the second session a second go-nogo task was presented instead. For the false-belief tasks, four different scenarios were used, each of which could be told as a prediction or explanation version. One of all possible sequences of these four scenarios was randomly assigned to each child. Similarly, one of the possible sequences of the three versions of the go-nogo task was randomly assigned to each child.

Procedure and Materials

Card Sorting

The procedure and materials were the same as in Experiment 1 with the following exceptions. The stimulus dimensions were color and number. Each of the target cards (glued to the posting box) showed either two green cars or one yellow car, and the test cards (to be sorted) showed either two yellow cars or one green car. The preswitch dimension was always color and the postswitch dimension was always number. There was a greater number of sorting trials than in Experiment 1: seven preswitch and nine postswitch.

False-Belief Tasks

Four traditional false-belief tasks were administered. Two scenarios consisted of stories enacted in a three-dimensional model (30 cm × 22 cm × 22 cm),

the other two consisted of picture stories. As in Experiment 1, there was a story about a dog looking for his unexpectedly transferred bone. A picture story involved a girl looking for her chocolate bar that was unexpectedly transferred. Two further stories were about children looking for a bird (enacted) or a rabbit (picture), who had moved unexpectedly to a new location.

False-belief understanding in the prediction version of each story was assessed by asking children where the protagonist would first go to get the object (or animal). The memory of critical story events was checked by three control questions: (1) where the object was now, (2) where the object was in the beginning, and (3) how the object got to the new location.

In the explanation version the children were told where the protagonist would look and were then asked to explain the protagonist's erroneous action by answering an open "why" question (i.e., analogous to the free answers to the open question in Experiment 1. In this traditional explanation task it was not possible to ask forced-choice questions as in the twin tasks). After giving their answer, children were asked one memory question about where the object was now.

Go-Nogo Tasks

For each task, participants were seated in front of a laptop on which one of two stimuli (6.5 × 6.5 cm), a red square or a yellow square, were presented one at a time in the center of the display (16 cm × 21.5 cm). Stimulus duration was 2 s for the 10 practice trials and 1 s for the 25 test trials. Time from stimulus offset to next stimulus onset (ISI) was .5 s for the 75–25 version (approximately 75% go and 25% nogo trials; i.e., precisely 19:6 trials) and for one of the 50–50 versions (precisely 12:13 trials). For the other 50–50 version the ISI was 1 s. For the practice trials in all three versions the ISI was 2 s. Children responded by pressing a light switch, which was affixed to a 23 cm × 10 cm × 10 cm wooden box.

The practice task began with the experimenter explaining the rules with reference to the light switch: "If something yellow appears, you press (Rule 1) and if something red appears, you do not press (Rule 2)." Before starting the practice task the experimenter checked whether the children had understood the rules. No child had any serious problems. After each practice stimulus children were given feedback about performance. Before starting the experimental task, the experimenter checked again the children's rule knowledge. During the experimental task no feedback was given.

Table 4   Number of Children Who Gave Correct Answers to Test Questions on Explanation Tasks and on Prediction Tasks in Experiment 2

Number of Correct Predictions	Number of Correct Explanations			Total
	0	1	2	
0	18	2	2	22
1	10	2	1	13
2	15	10	13	38
Total	43	14	16	73

Results and Discussion

Theory of Mind

False-Belief Prediction

Sixteen children gave wrong answers on at least one of the three control questions in one of the two stories. Performance on the control questions was partialed out, and it was found that the interpretation of the main results was not affected. Answers to the test questions again demonstrated a strongly bimodal distribution, as shown in the last column of Table 4. There was no indication that children gave fewer correct responses in the first session (60% correct) than in the second session (62%). Performance in the two test sessions showed good retest reliability, $\kappa = .63$.

False-Belief Explanation

Only 1 child gave any wrong answers to the control question in the two stories. Children's answers to the open test question were classified according to the same categories as in Experiment 1. In 31 cases, the children referred to a mental state (Category 1); and in 15 answers, they made note of relevant story facts (Category 2). In 29 cases, children referred to the protagonist's desire (Category 3), and in 32 cases they made reference to the desired object's location (Category 4). There were 19 irrelevant facts (Category 5) mentioned, and 20 "no" or "don't know" answers (Category 6). The categories were classified as "correct answers" (1 + 2) or "incorrect answers" (3–6) as in Experiment 1, following the procedure by Wimmer and Mayringer (1998).

As the last row in Table 4 shows, the majority of children failed to give any correct explanation. Fourteen children gave a correct explanation in only one of the sessions. There was some indication that fewer children gave correct answers in the first session (5 of the 14) than in the second session (9 of the 14). This trend, however, was not statistically reliable, Binomial Test: $N = 14, x = 5, p > .10$. Performance in the two test sessions yielded moderate retest reliability, $\kappa = .56$.

Relating Prediction and Explanation

The main body of Table 4 shows that 15 children gave two correct predictions without any correct explanations, whereas only 2 children gave two correct explanations but no correct predictions. Moreover, another 20 children gave one more correct prediction than correct explanation, whereas only 3 children gave one more correct explanation than correct prediction. Thus, predicting a belief-based action was clearly easier than explaining such an action in response to an open-ended "why?" question, Wilcoxon Signed Ranks test, $Z = -4.31, p < .001$.

Card Sorting (DCCS)

In the postswitch phase, the majority of children sorted either mostly correctly or incorrectly. When classified into three distinct categories of 0–2/3–6/7–9 correct, there was a slight but nonsignificant improvement from the first session (37/3/33) to the second session (33/2/38). The retest reliability using these three categories was quite good, $\kappa = .69$, as well as the correlation based on the continuous measure of items sorted correctly, $r = .76, p < .01$.

Go-Nogo

Children's performance on each task was scored in terms of percentage correct presses in response to go-stimuli (hits), and percentage correct inhibitions (correct rejections) in response to nogo-stimuli. (Also computed was the difference between hits and false alarms as a measure of accuracy. Because the data for this measure closely resembled the results for the hits, they are not reported.) To investigate age trends, children were grouped into four equal-size age groups (see Table 5); the first three groups had 18 children each, and the last group had 19 children). Independent 3 × 4 analyses of variance were carried out for each of the two performance measures with the three task versions as a within-participant factor and the four age groups as a between-participants factor.

The analysis for percentage correct presses (hits) showed a significant age trend only, $F(3, 69) = 14.22, p < .001$. This was due to the fact that there was an increase in percentage hits with age; in particular, from the youngest group with 37.5% to the next oldest with 60%, Least significant difference: $p < .001$, followed

Table 5   Percentage Correct Inhibitions on Go-Nogo Tasks in Experiment 2

Tasks (ISI)	Age Groups				Difference between Oldest and Youngest
	2,9–3,5	3,6–3,10	3,11–4,2	4,3–5,8	
50–50 (1 s)	87.2	89.7	89.3	94.7	7.5
50–50 (.5 s)	89.3	85.9	90.2	92.7	3.4
75–25 (.5 s)	71.3	78.7	89.8	94.7	23.4
Difference (ISI = .5 s)					
50–50 − 75–25	18.0	7.2	.4	−2.0	

by minimal nonsignificant further improvements to 69.9% and 66.6% in the two oldest groups.

The analysis of correct inhibitions showed significant main effects for age, $F(3, 69) = 3.01$, $p < .04$, and task, $F(2, 138) = 5.52$, $p < .01$, as well as for their interaction, $F(6, 138) = 2.71$, $p < .02$. The means in Table 5 show that the significant interaction was due to the fact that improvement with age was considerably stronger on the 75–25 task than on the two 50–50 tasks. The last column in Table 5 highlights this by reporting the difference between youngest and oldest group for the three different tasks. The last row shows that only the youngest group and, to some degree, the second to youngest group showed noticeably less inhibition on the 75–25 task than on the other tasks.

These results confirmed the reports of earlier studies (Beiswenger, 1968; Jarvis, 1968; Livesey & Morgan, 1991; Luria & Tizard, 1961; Miller et al., 1970) with the go-nogo task that there is a significant im-

provement on this task in the age range of 3 to 5 years. This allowed for the ability to check whether the improvement in inhibitory abilities was related to the other tasks and whether the relation between false-belief understanding and card sorting was due to changes in inhibitory capacities.

Relations between All Tasks

Table 6 shows the mean percentage correct and standard deviations for each variable of interest and the correlations between these variables. Experiment 2 replicated the finding from Experiment 1 that mastery of the false-belief task related to performance on the card-sorting task and the explanation version did so as much as, if not more than, the prediction version. When age, verbal intelligence, and performance on control questions were partialed out, the explanation task still related significantly to card sorting.

In general, the correlations tended to be somewhat lower in this experiment than in Experiment 1. One speculative reason for this could be the longer test sessions. In an informal comparison of several studies in this area, Perner and Lang (1999) found a negative correlation between length of individual testing sessions and correlations between tasks reported.

An important question motivating this experiment was to what degree the correlation between false-belief tasks and card sorting was due to the ability to inhibit a prevalent response strategy as measured by the go-nogo task. The last column in Table 6 indicates that the ability to inhibit responses to the nogo stimulus correlated significantly only with false-belief explanation, but not with the prediction or the card-sorting tasks. Moreover, partialing out correct inhibi-

Table 6   Basic Statistics and Correlations

Variables	Mean % (SD)	BPVS	False Belief			Card Sorting	Go-Nogo	
			Control Question	Prediction	Explanation		Hits	Correct Inhibition
Age		.54**	.38**	.60**	.53**	.45**	.53**	.34**
BPVS	—		.32**	.33**	.52**	.49**	.48**	.42**
Control question	94.9 (12.1)	—		.41**	.28*	.35**	.30**	.15
FB prediction	61.0 (44.3)	—	—		.38**	.39**	.31**	.15
FB explanation	31.5 (41.2)	—	—	[.08]		.60**	.43**	.32**
Card sorting	50.4 (44.7)	—	—	[.14]	[.42**]		.46**	.20
Go-nogo								
Hits	58.6 (20.5)	—	—	[−.04]	[.13]	[.21]		.40**
Correct inhibition	87.9 (13.2)	—	—	[−.08]	[.09]	[−.04]	[.21]	

Note: Correlations in brackets below the main diagonal are after age, BPVS, and control questions were partialed out. FB = false belief; BPVS = British Picture Vocabulary Scale.
* $p < .05$; ** $p < .01$.

tion performance on the go-nogo task affected the correlations between the other variables only marginally by at most .02 correlation points. Clearly, the observed correlations between belief task and card sorting were not due to the developing ability to inhibit strong response tendencies as measured by the go-nogo task.

GENERAL DISCUSSION

The main finding of this set of experiments was the strong correlation between the explanation version of the false-belief task and card sorting as a paradigm example of a task requiring executive control. Both experiments confirmed that there was a strong and robust correlation (in Experiment 2 the robustness was limited to the explanation version). Experiment 2 made clear that this correlation could not be due to developmental changes in the ability to inhibit unwanted responses as measured by the go-nogo task. The fact that in both experiments the explanation task was in no way easier than the prediction task, and the fact that the explanation task correlated with card sorting as strongly as the prediction task (as Russell et al., 1991, explicitly suggested) showed that it was not the executive demand of the prediction task of having to inhibit salient reality or salient thoughts about the desired object that was solely responsible for the correlation with executive tasks such as card sorting. The possibility that the explanation task, too, might have been susceptible to wrong answers in terms of reference to the salient desired object was ruled out by the forced-choice question in the identical-looking twins story of Experiment 1, in which the data clearly showed that children did not have any systematic tendency to answer wrongly. Rather they either were consistently correct (those who had made correct predictions) or they randomly guessed one of the choices (those that consistently failed prediction).

This finding suggests that there was no strong inhibitory problem for children in the false-belief task that would be solely responsible for the observed correlations with executive function tasks. This conclusion can be confirmed by several other findings in the literature that show that task variations in which the alleged inhibitory problem is reduced are not discernibly easier for children. For instance, Russell (1996) reported that 3-year-olds still have severe problems with a version of the false-belief task in which five locations are used and the children are merely told that the object is being relocated from its original box to one of the other four without actually seeing into which one. This information should greatly help suppress the automatic tendency to point to where the

object is, because the children don't actually know where that is. Unfortunately no direct comparison with the standard version was made. Robinson and Beck (2000, Experiment 4) did contrast this version with the standard version (although in a counterfactual reasoning task) and found no difference.

Finally, Clements and Perner (1994) found that a majority of children as young as 2,11 to 3,2 indicate some understanding of false belief by looking to the location where the protagonist mistakenly thinks the object is, a good year before they use this location in their answers to the question regarding where the protagonist will look for the object. If children's problem were one of explicitly inhibiting a prepotent response or idea (salient reality), then it should be more difficult to achieve this inhibition for implicit, unconscious-looking responses than for explicit, conscious answers to questions.

Therefore, we conclude that, contrary to the claim by Russell et al. (1991) and Hughes and Russell (1993), inhibition of a salient idea (reality), disengaging attention from the desired object, or inhibiting a prepotent response are not the main or only problem for children in the false-belief task. Consequently the observed developmental correlation between theory-of-mind and executive tasks requires a different explanation. In other words, by weakening the first and most plausible explanation (Russell et al., 1991) of the observed correlation between theory-of-mind tasks and executive function tasks, we reveal the need for other explanations. There are already several attempts in the literature to explain this relation; for example, a deeper underlying change that affects both kinds of abilities or the fact that one kind of ability is a prerequisite for the other ability. Below is a presentation of these theories in chronological order and, where applicable, mention of possible problems raised by the present data.

1. Wimmer (1989; Perner, 1991) suggested that children gain better self-control with a better understanding of their mind. Frith (1992) used this argument to explain the coincidence of theory-of-mind problems and inhibition problems in schizophrenic patients, and Carruthers (1996) used it to explain the same coincidence in children with autism. The present study's data do not speak against this position in general. They do, however, pose a problem for the more specific theory of how false-belief understanding relates to "executive inhibition" formulated by Perner, Stummer, and Lang (1999). Both false-belief understanding and inhibition of unwanted, interfering action schemata require an understanding of the causal effects of mental states; that is, that the false-belief task makes people look in the wrong place and a pre-

potent idea or action tendency makes people act against their better intentions. The false-belief task indexes the development of this understanding, which then helps children to exert better executive control over their unwanted interfering action tendencies. This explanation was thought to apply to the card-sorting as well as to the go-nogo task. The present finding that performance on the go-nogo task was not substantially related to the link between false-belief understanding and card sorting speaks against this particular application of the more general idea that a better understanding of the mind leads to better self-control.

2. Ozonoff, Pennington, and Rogers (1991) suggested that theory-of-mind and executive abilities are both impaired in children with autism because these abilities are served by the same brain region, which might be damaged or malfunctioning in this population. This suggestion can be extrapolated to normal development by assuming that maturation of this region accounts for the co-occurrence.

3. Frye et al. (1995) proposed that the ability to reason with embedded conditionals (if–if–then) accounts for the developmental correlation, because the relevant theory-of-mind and inhibition tasks all require such a reasoning structure. The finding in the present study that children's inhibition problems in the go-nogo task were few and that there was no substantial correlation with card sorting fits this theory, because the go-nogo task has a simple if–then structure, $<a_1 \rightarrow$ press, $a_2 \rightarrow$ do not press$>$, and should therefore be mastered well before the card-sorting task with its more complicated if–if–then structure: $<$Setting 1: $a_1 \rightarrow c_1$, $a_2 \rightarrow c_2$; Setting 2: $a_1 \rightarrow c_2$, $a_2 \rightarrow c_1>$.

However, the data highlight one of the problems of applying this structural analysis to the belief task pointed out by Perner and Lang (1999). In particular, Frye, Zelazo, Brooks, and Samuels (1996) have shown that the structurally analogous "ramp task" becomes much easier when only one of the antecedents is used, and Lang (1999) found the same for the card-sorting task (i.e., only one kind of test card used). The analysis of the false-belief task provided by Frye et al. (1995) only mentions one antecedent (a_1: "when looking for the chocolate") that leads, under one setting condition (own perspective), to one action (c_1: go to object's actual place); but under the other setting condition (story character's perspective) leads to another action (c_2: go to object's original place). Under this structural description, the false-belief task is a "single-antecedent task" and should be much easier than the card-sorting task (a two-antecedents task). The present study's data did not provide any support for this implication: the false-belief task was as difficult as card sorting.

4. Finally, Russell (1996, 1998, p. 295) argued that "the monitoring of actions and the ability to act at will are necessary ingredients to the development of a 'pretheoretical' form of self-awareness [and] . . . that this form of self-awareness must be in place if the individual is to gain an adequate grasp of mental concepts." Although Russell and Hill (2001) rejected the action-monitoring-deficit hypothesis for autism, Pacherie (1998) saw a particular problem for children with autism in forming motor images (Jeannerod, 1997), which are formed when the execution of motor programs is delayed or blocked and play a role in trouble shooting and error correction. Motor images also make the intentions in action conscious and could be responsible for the executive problems mastered around 4 years and provide the building blocks for the theory-of-mind development at this age.

The data presented in this article do not, strictly speaking, rule out any of these newer theories. The data do, however, speak against the original, most plausible explanation that the observed relations between theory-of-mind and executive function tasks are due to problems of inhibition in the theory-of-mind tasks. With this option made implausible, future research can become fully focused on testing the newer and more substantive theories about the relation between theory-of-mind development and executive control.

The data also contribute to the understanding of which aspects of executive tasks are responsible for the developmental link with understanding false belief. The finding that improvements on the go-nogo task tend to happen before the belief task is mastered and the fact that there was no substantive correlation between these tasks suggests that it is not inhibition of a response, as such, that constitutes the relevant feature. Similarly, Carlson and Moses (2001; Carlson, Moses, & Breton, in press) reported that "delay" tasks that require inhibition of a prepotent temptation like the go-nogo task did not correlate very strongly with false-belief tasks. Only "conflict" tasks that required inhibition and some additional cognitive load (e.g., activate a novel response or make the nonresponding dependent on some background feature as in "Simon Says") correlated strongly, and did so even after the influence of age and intelligence (Hughes, 1998a) and working memory and performance on delay tasks (Carlson et al., in press) had been accounted for. Similarly in the present data, the go-nogo task showed little relation with false belief after age and intelligence had been partialed out, whereas card sorting remained significantly related to the false-belief task—in particular, the explanation version. The good correlation with the explanation version adds to this

picture the finding that the role of inhibition does not reside in a simple inhibitory component of the standard false-belief prediction task.

ACKNOWLEDGMENTS

ADDRESSES AND AFFILIATIONS

Corresponding author: Josef Perner, Department of Psychology, University of Salzburg, Hellbrunnerstr. 34, A-5020 Salzburg, Austria; e-mail: josef.perner@ sbg.ac.at. Birgit Lang and Daniela Kloo are also at the University of Salzburg.

APPENDIX: FALSE-BELIEF EXPLANATION (TWIN TASK)

Episode 1

Two identical-looking twins are playing ball in a room. Then they place the ball into one of two boxes and one of them leaves the scene because he wants some milk.

Check question: "Where is the ball?"

Episode 2

The other twin takes the ball out of the box and again plays with the ball for awhile. Then he places the ball into the other box and leaves the room as well.

Check questions:

1. "Where is the ball now?"
2. "Has the other one, who had gone outside in order to drink some milk, seen that the ball is now in the other box?"

Episode 3

Both twins return with their mother who asks them where their ball is. One twin goes to where the ball actually is, the other one goes to the other box.

Test questions about the mistaken twin:

1. Open question: "Why does this one look in this box?"

2. Forced-choice question: "Is it because he had stayed in the room and played with the ball or is it because he had left the room in order to drink some milk in the kitchen?"

REFERENCES

Astington, J. W. (1996). What is theoretical about the child's theory of mind?: Vygotskian view of its development. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 184–199). New York: Cambridge University Press.

Bartsch, K., & Wellman, H. M. (1989). Young children's attribution of action to beliefs and desires. *Child Development, 60*, 946–964.

Beiswenger, H. (1968). Luria's model of the verbal control of behavior. *Merrill Palmer Quarterly, 14*, 267–284.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Carlson, S. M., Moses, L. J. & Breton, C. (in press). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*.

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

Carruthers, P. (1996). *Language, thought and consciousness. An essay in philosophical psychology.* Cambridge, U.K.: Cambridge University Press.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377–397.

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47*, 25–31.

Doherty, M J., & Perner, J. (1998). Metalinguistic awareness and theory of mind: Just two words for the same thing? *Cognitive Development, 13*, 279–305.

Dunn, L. M., Dunn, J., Whetton, C., & Pintillie, D. (1982). *British picture vocabulary scale (BPVS)*. London: NFER–NELSON.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance–reality distinction. *Cognitive Psychology, 15*, 95–120.

Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition, 44*, 283–296.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hillsdale, NJ: Erlbaum.

Frye, D., Zelazo, P. D., Brooks, P. J., & Samuels, M. C. (1996). Inference and action in early causal reasoning. *Developmental Psychology, 32*, 120–131.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483–527.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Gordon, A. C., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology, 68*, 70–83.

Harris, P. L. (1993). Pretending and planning. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 283–304). Oxford, UK: Oxford University Press.

Hughes, C. (1998a). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*, 233–253.

Hughes, C. (1998b). Finding your marbles: Does preschoolers' strategic behavior predict later understanding? *Developmental Psychology, 34*, 1326–1339.

Hughes, C., Dunn, J., & White, A. (1998). Trick or treat?: Uneven understanding of mind and emotion and executive dysfunction in 'hard to manage' preschoolers. *Journal of Child Psychology and Psychiatry, 39*, 981–994.

Hughes, C., & Russell, J. (1993). Autistic children's difficulty with mental disengagement from an object: Its implication for theories of autism. *Developmental Psychology, 29*, 498–510.

Jarvis, P.E. (1968). Verbal control of sensory-motor performance: A test of Luria's hypothesis. *Human Development, 11*, 172–183.

Jeannerod, M. (1997). *The cognitive neuroscience in action*. Oxford, U.K.: Blackwell.

Keenan, T. (1999, April). *The role of working memory and inhibitory control in children's understanding of false belief*. Paper presented at the biennial meeting of the Society for Research in Child Development, Albuquerque, New Mexico.

Lang, B. (1999). [Sorting with only one kind of card in the DCCS task]. Unpublished raw data.

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science, 1*, 247–253.

Livesey, D. J., & Morgan, G. A. (1991). The development of response inhibition in 4- and 5-year-old children. *Australian Journal of Psychology, 43*, 133–137.

Luria, A., & Tizard, J. (Eds.). (1961). *The role of speech in the regulation of normal and abnormal behaviour*. New York: Pergamon.

Melchers, P., & Preuß, U. (1991). *Kaufman Assessment Battery for Children* (German version). Frankfurt am Main, Germany: Swets & Zeitlinger.

Miller, S. A., Shelton, J., & Flavell, J. H. (1970). A test of Luria's hypothesis concerning the development of verbal self-regulation. *Child Development, 41*, 651–665.

Mitchell, P. (1996). *Acquiring a conception of mind: A review of psychological research and theory*. East Sussex, U.K.: Psychology Press.

Mitchell, P., & Lacohée, R. (1991). Children's early understanding of false belief. *Cognition, 39*, 107–127.

Moses, L. J., & Flavell, J. H. (1990). Inferring false beliefs from actions and reactions. *Child Development, 61*, 929–945.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic children: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32*, 1081–1105.

Pacherie, E. (1998). Motor-images, self consciousness and autism. In J. Russell (Ed.), *Autism as an executive disorder* (pp. 215–255). Oxford, U.K.: Oxford University Press.

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

Perner, J. (1995). The many faces of belief: Reflections on Fodor's and the child's theory of mind. *Cognition, 57*, 241–269.

Perner, J., Baker, S., & Hutton, D. (1994). Prelief: The conceptual origins of belief and pretence. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development.* Hove, U.K.: Erlbaum.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3*, 337–344.

Perner, J., & Lang, B. (2000). Theory of mind and executive function: Is there a developmental relationship? In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from autism and developmental cognitive neuroscience* (pp. 150–181). Oxford, U.K.: Oxford University Press.

Perner, J., Stummer, S., & Lang, B. (1999). Executive functions and theory of mind: Cognitive complexity or functional dependence? In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self-control* (pp. 133–152). Mahwah, NJ: Erlbaum.

Robinson, E J., & Beck, S. R. (2000). What is difficult about counterfactual reasoning? In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 101–119). Hove, U.K.: Psychology Press.

Robinson, E. J., & Mitchell, P. (1995). Masking of children's early understanding of the representational mind: Backwards explanation versus prediction. *Child Development, 66*, 1022–1039.

Russell, J. (1996). *Agency. Its role in mental development.* Hove, U.K.: Erlbaum Taylor & Francis.

Russell, J. (1998). How executive disorders can bring about an inadequate "theory of mind." In J. Russell (Ed.), *Autism as an executive disorder* (pp. 256–299). Oxford, U.K.: Oxford University Press.

Russell, J., & Hill, E. L. (2001). Action-monitoring and intention reporting in children with autism. *Journal of Child Psychology and Psychiatry, 42*, 317–328.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The "windows task" as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology, 9*, 331–349.

Shrout, P. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research, 7*, 301–317.

Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition, 30*, 239–277.

Wimmer, H. (1989). Common-Sense Mentalismus und Emotion: Entwicklungspsychologische Implikationen [Common-sense mentalism and emotion: Implications

for development psychology]. I. In E. Roth (Ed.), *Denken und Fühlen* (pp. 56–66). Berlin: Springer.

Wimmer, H., & Mayringer, H. (1998). False belief understanding in young children: Explanations do not develop before predictions. *International Journal of Behavioral Development, 22*, 403–422.

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.

Wimmer, H., & Weichbold, V. (1994). Children's theory of mind: Fodor's heuristics examined. *Cognition, 53*, 45–57.

Zaitchik, D. (1991). Is only seeing really believing?: Sources of the true belief in the false belief task. *Cognitive Development, 6*, 91–103.