Training Transfer Between Card Sorting and False Belief Understanding: Helping Children Apply Conflicting Descriptions

Daniela Kloo and Josef Perner

Two studies investigated the parallel developmental progress in theory of mind and executive control, as exemplified by correlations between the Dimensional Change Card Sorting task (DCCS; Frye, Zelazo, & Palfai, 1995) and the false-belief task. Experiment 1 with sixty 3-year-old children confirmed earlier studies (e.g., Perner & Lang, 2002), suggesting that children’s problem with the DCCS lies in the redescription of stimulus objects. Experiment 2 with forty-four 3- to 4-year-olds reinforced the correlational evidence for a developmental link by showing transfer of training: False-belief training improved DCCS performance, and DCCS training significantly increased children’s performance on the false-belief task. Results are discussed in relation to 5 theories explaining the existing correlational evidence.

At around 4 years of age, important developmental changes occur. At this age, children develop an understanding that people may have different mental representations of the world and that one can be mistaken about the world. A widely used measure designed to tap these crucial changes in children’s theory of mind is the false-belief task (Wimmer & Perner, 1983). At about the same age, children also improve markedly in a host of executive-function tests, which are measures of children’s growing self-control. There is increasing evidence that this parallel development is not merely coincidental. Several recent studies have demonstrated correlations between developmental advances in theory of mind and executive function in the age range of 3 to 5 years (e.g., Carlson & Moses, 2001; Hughes, 1998; Perner, Lang, & Kloo, 2002; Russell, Mauthner, Sharpe, & Tidswell, 1991; for a review, see Perner & Lang, 1999).

Competing theories aim at explaining this observed link between theory of mind and executive functions. Three theories suggest that performance on theory-of-mind tasks and executive-function tasks is determined by a single process. The first explanation (Hughes & Russell, 1993; Russell et al., 1991) was that typical theory-of-mind tasks contain executive components and, therefore, correlate with executive tasks. The claim is that in the false-belief task children need to inhibit their salient knowledge about the current location of the object to predict the protagonist’s false-belief-based action. Some empirical support for this position was provided, for example, by studies showing that children perform better on strategic deception tasks when the inhibitory demands are reduced (Carlson, Moses, & Hix, 1998; Hala & Russell, 2001). However, at least one clear prediction from this theory (Russell et al., 1991, p. 342) for the false-belief task has been disconfirmed, namely, that the explanation version of the false-belief task, which does not share the alleged executive problem, should be easier than its prediction version. The explanation version, when controlling for correct guesses, is not easier than the prediction version and correlates with executive tasks as strongly as the prediction version (Hughes, 1998; Perner, Lang, et al., 2002).

Another explanation for the observed developmental relationship is that theory of mind and executive functions may be served by the same
region of the frontal cortex (e.g., Ozonoff, Pennington, & Rogers, 1991). This could explain the observed association of theory of mind and executive problems in autism. Later development of these structures can then also explain why executive-function and theory-of-mind performance are related in normal development. According to recent imaging studies, theory-of-mind tasks tend to activate anterior parts of the anterior cingulate cortex (ACC). However, only some executive tasks, namely, Stroop-like tasks that have an inhibitory component, are associated with the ACC, though most of these activations lie posterior to the region activated by theory-of-mind tasks. Other classical executive tasks, especially working-memory and problem-solving tasks, primarily activate the dorsolateral prefrontal cortex (for recent reviews, see Frith & Frith, 2003; Kain & Perner, 2003). Nevertheless, the maturational schedule of the entire ACC could explain the developmental link between theory-of-mind tasks and Stroop-like executive tasks.

Zelazo and Frye (1997) argued that the correlation between theory of mind and executive function can be explained in terms of general reasoning abilities. According to their cognitive complexity and control (CCC) theory, rule-use tasks, such as the Dimensional Change Card Sorting task (DCCS task; Frye, Zelazo, & Palfai, 1995), and theory-of-mind tasks require a common logical structure, namely, reasoning with embedded conditionals of the form “if condition then if antecedent then consequent.” This theory is discussed in more detail in relation to Experiment 1.

Two further theories envisage a functional dependency between the development of executive control and theory of mind. For example, Wimmer (1989) and Perner (1991, 1998) suggested that improved understanding of one’s mind leads to greater self-control as the child develops; that is, theory of mind is seen as a prerequisite for executive functions. In contrast, Russell (1996) argued that self-control is a prerequisite for building a theory of mind because theory of mind is grounded on first-person experience of agency.

However, the specific nature of the relationship between theory of mind and executive function is far from clear. One problem is that executive function is an umbrella term “so broad as to lose meaning” (Towse, Redbond, Houston-Price, & Cook, 2000, pp. 348–349). There is no consensus of what behaviors indicate executive control. In a factor analysis of typical executive-function tasks, Pennington (1997) identified at least three distinguishable dimensions: (a) inhibition, (b) cognitive flexibility or set shifting, and (c) working memory. Moreover, different factors or processes are often mixed in tasks, and it remains to be determined exactly what makes a task difficult. For example, the DCCS task (Frye et al., 1995) has repeatedly been shown to be specifically correlated with the false-belief task even when age and verbal intelligence are partialled out (e.g., Carlson & Moses, 2001; Frye et al., 1995; Perner, Lang, et al., 2002). However, the source of children’s difficulty with this sorting task is still under debate. Therefore, the nature of children’s problems in this executive task was investigated in Experiment 1.

A second problem is that there are only correlational studies showing a relationship between theory of mind and executive function. An outstanding question is whether it is possible to substantiate further this relationship using another research design, for example, a training study. Does theory-of-mind training improve performance on executive tasks such as the DCCS and vice versa? Such a training study could also shed light on the different theories that aim at explaining the developmental relationship, and was therefore undertaken in Experiment 2.

**EXPERIMENT 1**

The DCCS task (Frye et al., 1995) is frequently used to assess young children’s executive abilities. In the standard version, children are required to sort first according to one dimension and second according to another dimension. Two target cards, each affixed to one of two sorting boxes, are used. They depict, for example, a red apple and a blue pear. The test cards (red pears and blue apples) match one target card on one dimension and the other target card on the other dimension. In the preswitch phase, children are told a pair of rules, for example, the color rules: They are asked to sort all the blue cards into the box portraying something blue and to sort all the red cards into the box displaying something red. Typically, 3-year-olds have no problems when sorting the cards according to one dimension. However, they usually have difficulties in the postswitch phase when the sorting rules change: Now the cards should be sorted according to the other dimension (e.g., according to shape). After this switch, it is not until about the age of 4 years that children continue to sort correctly.

Several explanations have been proposed in the literature to explain 3-year-olds’ difficulty. The traditional explanation (e.g., Zelazo & Frye, 1997) referred to CCC theory. Complexity is measured in terms of the number of levels of embedding inherent
in complex rule systems (systems of condition–
action statements). According to CCC theory, 3-year-
olds cannot represent “a higher-order rule that
integrates two incompatible pairs of rules” (Zelazo
& Frye, 1997, p. 120). Younger children are thought
unable to employ an embedded if-if-then rule
structure to use such a higher order rule (e.g., “If
we are playing the color game, then if I give you a
red pear, then put it into the box with the red apple
target) for selecting the appropriate consequence.

However, some recent findings speak against this
account. Children of 3 years easily master a card-
sorting task if no target cards are used, although
the structure of embedded incompatible rules remains
the same (Perner & Lang, 2002; Towse et al., 2000).
Children are able to switch their responses to an
incompatible pair of rules in a reversal shift task in
which rules change solely within one dimension
(Brooks, Hanauer, Padowska, & Rosman, 2003,
Experiment 1; Perner & Lang, 2002). Furthermore,
children’s performance in the standard task improve
when they are induced to redescribe the test
cards according to the postswitch dimension,
although the if-if-then structure of embedded con-
flicting rules is still present (Kirkham, Cruess, &
Diamond, in press; Towse et al., 2000).

For example, Towse et al. (2000, Experiment 4)
asked children who had made an error in the
postswitch phase to identify a test card (“What is
this?”). Children who described this test card accord-
ing to the preswitch dimension were asked to label
the card according to the postswitch dimension (e.g.,
“Is it a red or green card?”). Then, the children were
asked to sort the card. About one third of the
children (7 of 22) recovered from erroneous sorting
in the DCCS task when they were prompted to label
the test card by the new dimension. Similarly,
Kirkham et al. (in press) found that if children are
instructed to label the test card by the relevant
dimension before sorting it, they err less frequently.

Another explanation of children’s difficulty with
the DCCS task is failure to exert executive inhibition
of an interfering response tendency at the level of
action schemas (Carlson & Moses, 2001; Perner,
Stummer, & Lang, 1999). According to this explana-
tion, correct sorting in the postswitch phase requires
inhibition of the interfering action schema practiced
in the preswitch phase. In fact, initial findings (e.g.,
Zelazo, Frye, & Rapus, 1996) spoke in favor of
inhibition problems purely at the level of action
schemas because children were able to answer
knowledge questions (e.g., “Where do the red
things go in the shape game?”) about the sorting rules
correctly while sorting the cards incorrectly. But at
least two findings speak against inhibition problems
limited to the level of actions. Munakata and Yerys
(2001) showed that the dissociations between knowl-
edge and action disappear when knowledge ques-
tions and sorting measures are more closely matched
for amount of conflict by using knowledge questions
mentioning both dimensions (e.g., “Where do the red
trucks go in the shape game?”). Perhaps the most
convincing evidence against inhibition problems at
the level of action schemas comes from Jacques,
Zelazo, Kirkham, and Semcsesen (1999). They found
that 3-year-olds’ problems in the DCCS task persist
even when they do not have to sort themselves but
have to judge a puppet’s sorting.

An alternative explanation for 3-year-olds’ diffi-
culty with the DCCS task (see Perner & Lang, 2002)
is that children find it difficult to redescribe the
objects on the cards because the switch in rule is a
switch in dimension. For example, if children must
describe the depicted objects on the cards as objects
of a certain color in the preswitch phase, they must
switch to describing them as objects with a certain
shape in the postswitch phase. Perner and Lang
(2002) found that children had no serious problems
if the switch in dimension (and consequently the
need to redescribe the cards) was avoided by using a
reversal shift instead of an extradimensional shift or
if the visual clash was avoided by replacing the
target cards by target characters. Therefore, the
combination of an extradimensional shift and a
visual clash between target and test cards seems to
be critical for the difficulty of the DCCS task. One
objective of Experiment 1 was to confirm this
finding. In Perner and Lang’s reversal shift condi-
tion, cards varied only with respect to the shape
dimension. Children had to sort, for example, the
cars to the car target and the suns to the sun target in
the preswitch phase (normal shape game). In the
postswitch phase, they had to sort all the cars to the
sun target and all the suns to the car target (reversed
shape game). In the target characters condition, the
target cards were replaced by a pair of target
pictures displaying familiar characters such as
Donald Duck. The switch from the preswitch
dimension to the postswitch dimension was de-
scribed as a change of the character’s preference. For
example, in the preswitch phase, Donald wanted all
red things and was to be given red cars, whereas in
the postswitch phase, he wanted all the suns and
consequently was to be given yellow suns.

Perner and Lang (2002) also found an order effect:
The standard DCCS was only difficult when
presented as the first task but not when given after
a nonstandard task. However, these easier versions
had no training effect in a similar experiment (Lang, 2001, Experiment 4), which confirmed nevertheless that the combination of an extradimensional shift and a visual clash is crucial for children's difficulties.

One objective of our first experiment, thus, was to clarify whether there is or is no such immediate training effect of these easier card-sorting versions (reversal shift and target characters version) on the standard DCCS task. A second objective was to confirm that these modifications of the DCCS task eliminate the difficulties of the standard task to feel more confident that children have difficulties because they must redescribe the cards. A third objective was to substantiate further the finding that performance on the standard DCCS task is related to mastery of the false-belief task.

Method

Participants

Sixty 3-year-old Caucasian children (31 girls and 29 boys) from nine nursery schools in Upper Austria and one in the city of Salzburg participated in the study. Children were predominantly from middle-class backgrounds. Their ages ranged from 2 years 11 months to 4 years 0 months (M age = 3, 8, SD = 3.80 months).

Design

Each child was tested individually in one session lasting about 15 min. There were three experimental groups: One group of children received two false-belief prediction tasks separated by a standard DCCS task. The other two groups were given first a false-belief prediction task then a nonstandard DCCS task (either a reversal shift task or a task using target characters) followed by a second false-belief prediction test and the standard DCCS. In the target characters version, each target box was marked either by a picture of a boy or of a girl. Test cards consisted of six green horses and six yellow fish. Children were asked to name the characters in the pictures. Then, the procedure was the same as in the standard task except that in the preswitch phase the girl was described as wanting all green things and the boy was described as wanting all yellow things. In the postswitch phase, the switch to the new dimension was explained as a change of preference (e.g., “The girl now wants all the fish.”). This resulted in a reversal of preferred cards. For example, the girl who wanted green things and got green horses now gets yellow fish because she wants fish.

In the reversal shift task, the 2 target and 12 test cards all had the same color (yellow) and differed only in shape (fish or horses). The procedure was the same as in the standard task except that in the preswitch phase children were asked to play the...
“correct” animals game, “All the horses go to the horse. And all the fish go to the fish.” In the postswitch phase, they were instructed to play the “wrong” animals game, “Now, all the horses go to the fish. And all the fish go to the horse.”

False Belief

Two traditional false-belief tasks (Wimmer & Perner, 1983) were administered. The first story involved a girl looking for a bird who had unexpectedly moved to another place and was enacted in a three-dimensional model (30 cm × 22 cm × 22 cm). The second was a picture story about a boy looking for his rabbit who had unexpectedly moved to a new location. The order of the two stories was counterbalanced across children.

In each story children had to predict the protagonist’s action based on the protagonist’s false belief: “Where will [for example] Andreas go first?” Memory of critical story events was checked by four control questions, for example: (a) “Where is the rabbit now?” (b) “Where did the boy put his rabbit at the beginning?” (c) “How did the rabbit get there (point)?” (d) “Could the boy see this?”

Results

Card Sorting

On each card-sorting task, children were given a score between 0 and 5 depending on the number of cards sorted correctly in the postswitch phase. First, we analyzed children’s performance on the three card-sorting versions (when given as the first task). A one-way analysis of variance (ANOVA) with task version as between-participants factor and postswitch performance as dependent measure revealed a significant effect of task version, F(2, 57) = 3.73, p < .05. Planned contrasts showed that the standard DCCS was significantly more difficult than the reversal shift task (p < .05) and the target characters version (p < .05). The two nonstandard tasks did not differ. The relevant means are shown in Table 1.

As a second step, we compared performance on the standard DCCS in the three experimental conditions: presented first, administered after a reversal shift, or given after a target characters task. A two-way ANOVA with the three experimental conditions and direction of shift (from color to shape vs. from shape to color) as between-participants factors and DCCS performance as dependent measure showed no significant effect. This indicates that practice on an easier task version had no training effect on performance in the standard DCCS and that direction of shift did not influence performance.

Relationship Between Card Sorting and False Belief

Because of the significant difference between performance on the first and second false-belief task, analysis was carried out separately for each test question. There was no significant relationship between the first test question and the DCCS task, r = .10. However, DCCS performance was correlated with the second false-belief question, r = .37, p < .01, Cohen’s d = .80. This relationship remained significant when age and memory of critical story events were partialed out, r = .28, p < .05, Cohen’s d = .58. Furthermore, of the children (n = 40) who had received one of the nonstandard tasks, the relation between the second false-belief question and the DCCS task remained significant when performance on the nonstandard task was partialed out, r = .35, p < .05, Cohen’s d = .75. There were no significant correlations between the false-belief tasks and the easier card-sorting tasks, −.12 ≤ r ≤ .33.

Discussion

Replicating previous studies (e.g., Carlson & Moses, 2001; Frye et al., 1995; Perner, Lang, et al., 2002), performance on the card-sorting task was correlated
with the ability to pass the second false-belief task. However, there was no significant correlation between the first false-belief task and the card-sorting task. One explanation for this finding, which could also explain the significant improvement on the second false-belief task, may be that some children did not show their true competence on the first false-belief task because this was always the first task in the session and children were not yet familiar with the experimenter or the testing procedure.

Replicating Perner and Lang (2002) as well as Lang (2001, Experiment 4), nonstandard card-sorting tasks either using target characters instead of target cards or including a reversal shift instead of an extradimensional shift were much easier than the traditional DCCS task. In contrast to Perner and Lang but in line with Lang (Experiment 4), these easier task versions exerted no significant training effect on the standard version, although children performed slightly better on the standard DCCS (69% correct) after receiving the target characters task than when the standard version was given first (61% correct).

The traditional explanation for children’s difficulty in the DCCS task (CCC theory) was that young children cannot comprehend the doubly embedded conditional if-if-then structure of the DCCS task (Zelazo & Frye, 1997). The other classical explanation (Carlson & Moses, 2001; Perner et al., 1999) focuses on younger children’s inability to inhibit learned-action tendencies. Once children have learned to put the blue apple with the blue pear in the color game, they cannot inhibit this action when they should put the blue apple with the red apple in the fruit game. Both these theories seem incompatible with three critical findings: (a) children of 3 years easily master a card-sorting task if no target cards are used (Experiment 1 reported here; Lang, 2001, Experiment 4; Perner & Lang, 2002), (b) children have no serious problems in a reversal-shift task in which rules change solely within one dimension (Experiment 1; Brooks et al., 2003, Experiment 1; Lang, 2001, Experiment 4; Perner & Lang, 2002), and (c) children’s difficulty declines when they are induced to redescribe the test cards according to the new dimension (Kirkham et al., in press; Towse et al., 2000). Nevertheless, in all these easier versions, children must still inhibit the preswitch action schema and employ an if-if-then structure of embedded incompatible rules.

To explain the finding that target cards are essential, Perner and Lang (2002) suggested that the use of target cards induces children to encode the sorting rules as the general rule “put each card to the corresponding target.” A recent experiment (Kloo, Perner, & Gaddy, 2002) supported this claim. In a card-sorting task using three test cards, children were given only two sorting rules (instead of three) but they had practically no problems sorting the third card according to the relevant dimension. This suggests that children do not apply pairs (or, in this case, triples) of rules in the DCCS task. In contrast, they seem to describe the cards according to one dimension and use a single, general rule (“put each card to the corresponding target”).

To put this rule into effect, children have to treat objects, for example, a blue apple as either a blue thing (in the color game) or as an apple (in the fruit game) but not both, or they would not know which box was the corresponding target. In the standard DCCS, during the instructions for the postswitch phase, children are only provided with verbal descriptions using the other dimension; for example, “Now the blue things go to the box with a blue thing on it” (as opposed to apples going to the box with an apple on it), but they are not explicitly told that the blue apple (so far treated as an apple) now has to be redescribed as a blue thing.

We suggest that children’s problem with the standard version is that they do not understand that redescription is possible, and consequently, they do not realize what the experimenter really meant with his or her instructions initiating the postswitch phase. These children are, however, better able to switch descriptions if explicitly prompted. This happens in the easy card-sorting tasks. When cards are sorted according to what some character wants (Experiment 1; Perner & Lang, 2002), it is explicitly said that one character switches from wanting apples to wanting red things. When children are explicitly asked to label the objects according to the new dimension (“Is it a red or green card?”; Towse et al., 2000, Experiment 4; Kirkham et al., in press), some switch. Or if children are shown for a single item that it is now, in the postswitch phase, to be put into the other box (Towse et al., 2000, Experiment 1), they realize that the object has to be treated as something different. Furthermore, eliminating the need to describe one and the same object differently (e.g., apple vs. blue thing) by separating the two dimensions displayed on the card (e.g., displaying an outline of an apple next to a patch of blue) improves performance considerably (Kloo & Perner, 2002).

EXPERIMENT 2

According to Experiment 1, mere exposure to single instances of easy task versions has no detectable
effect on a standard DCCS task administered in the same session. This confirms the statement by Frye et al. (1995, p. 505) that “the strength of the perseverative effect also suggests that training may not quickly change young children’s judgments in these types of tasks.” However, this raises the question whether a more intensive, explicit training might reliably improve children’s performance. Successful training studies of theory of mind (e.g., Clements, Rustin, & McCallum, 2000; Slaughter & Gopnik, 1996) emphasize that extensive explanations during training and sufficient time for the effects to settle are necessary for learning to occur.

Therefore, to enhance the chances for a training effect, we provided children with verbal feedback and explanations during the DCCS task in two training sessions over 2 weeks. Previous studies (Experiment 1; Kirkham et al., in press; Perner & Lang, 2002; Towsle et al., 2000) have indicated that children have difficulty in the postswitch phase because they have to describe the cards according to a new dimension. Therefore, we placed emphasis on the need to redescribe the cards after the switch with the relevant dimension. Another group of children was trained on false-belief tasks, and a control group was trained either on relative-clauses or on number-conservation tasks. These control training conditions were chosen because they have been used previously as control conditions in theory-of-mind training studies for this age group (e.g., Hale & Tager-Flusberg, 2003; Slaughter & Gopnik, 1996).

Training, in contrast to correlational studies, could help to differentiate between the two theories that see a functional relationship between theory of mind and executive functions. If theory of mind is a prerequisite for executive functions (e.g., Perner, 1991, 1998; Wimmer, 1989), one would not expect any training to improve executive control without some concomitant improvement in theory of mind as prerequisite for sustaining the increase in executive control. In contrast, if executive control is a prerequisite for theory of mind (e.g., Russell, 1996), any training should improve theory of mind only if it also induces the higher level of executive control prerequisite for the advance in theory of mind. A reciprocal transfer would be compatible with both theories.

Training effects would also speak against theories that see the reason for this developmental relation in the purely maturational development of frontal brain structures (e.g., Ozonoff et al., 1991) but would be compatible with a theory of stimulus-triggered brain development.

**Method**

**Participants**

Seventy-four Caucasian children (33 girls, 41 boys) between the ages of 3.0 and 4.7 (M = 46.5 months, SD = 4.7 months) from five nursery schools in predominantly middle-class areas in the city of Salzburg and in the suburbs were screened. All children who failed at least one of the two test questions of the false-belief task or sorted more than one card incorrectly in the postswitch phase of the DCCS were included in the study. Control task performance was not a criterion for inclusion.

Thirty children were disqualified for the following reasons: Nine had a standardized score below 80 in the verbal subtest of the Kaufman Assessment Battery for Children (K–ABC; Melchers & Preuß, 1991). Two children were excluded because they had difficulty with understanding basic task elements: One boy made more than one error in the preswitch phase of the DCCS, and 1 boy failed more than two (out of four) memory questions (see the following) of the false-belief task. Sixteen children were excluded because they performed too well on the DCCS task (sorting four or more of five cards correctly in the postswitch phase) and were perfect on the false-belief measures (passing both the false-belief prediction and explanation test questions). Three children were lost to illness or attrition.

The final sample consisted of 44 children (22 girls and 22 boys), ranging in age from 3.0 to 4.7 (M = 45.1 months, SD = 4.9 months). These children correctly answered 0.73 (SD = 0.58) of the two test questions of the false-belief task, sorted 1.59 (SD = 2.11) of 5 cards correctly in the postswitch phase of the DCCS, and responded correctly to 37.5% (SD = 40.2) of the test items of the control task. They were randomly allocated to one of three training groups matched for pretest performance. Finally, the cardsorting group consisted of 7 girls and 7 boys, the false-belief group consisted of 8 girls and 7 boys, and the control group consisted of 7 girls and 8 boys.

**Design**

First, the children were pretested with a false-belief task (including a prediction and an explanation test question), a card-sorting task, a verbal intelligence test (K–ABC), and a control task. Half of the children were randomly assigned to one of two control tasks: a relative-clause task (Nebensatzkonjunktion subtest from Penner, 1999) or a classic number-conservation task (Piaget, 1965).
Training started about 1 week \((M = 9.05 \text{ days}, SD = 3.14 \text{ days})\) after the pretest, followed by a second training session within approximately 1 week \((M = 7.93 \text{ days}, SD = 2.53 \text{ days})\) after the first. The posttest followed about 11 days \((M = 10.75 \text{ days}, SD = 4.29 \text{ days})\) after the second training session. The variation in timing sessions was due to illness and absence.

At posttest, children were retested on the DCCS task, the false-belief task, and the control task. These tasks used the same format as at pretest but differed in material. In addition, children received a novel version of the card-sorting task with three different test and target cards. The pretest and posttest sessions were conducted by a male experimenter who was blind to the training group membership of the children. Training was carried out by a female experimenter.

Children were tested and trained individually in a separate room. The pretest lasted approximately 20 min. Each training session lasted about 15 min. The posttest lasted approximately 30 min. After each session, children received stickers for “doing so well.”

Procedure and Materials

Pretest

Half of the children were given first the K–ABC test followed by a false-belief task, a control task, and a card-sorting task. The other half received the reverse order. As control task, half of the children received a relative-clause task; the other half received a number-conservation task.

Card sorting. Two sets of cards \((10 \text{ cm} \times 7 \text{ cm})\) were used. Each of these card sets consisted of two target cards (a big yellow horse and a small red fish; a big blue dog and a small green bird), which were each affixed on a \((22 \text{ cm} \times 15 \text{ cm} \times 14.5 \text{ cm})\) box. The test cards had to be placed into one of these boxes through a slit. There were 24 test cards: 6 with a small yellow horse, 6 with a big red fish, 6 with a small blue dog, and 6 with a big green bird. Each task involved a preswitch phase and a postswitch phase. The direction of shift was always from shape to size. Half of the children received a particular set at pretest. The other half were given this set at posttest.

The procedure was the same as in the standard DCCS administered in Experiment 1 with the following exceptions: At the beginning of the preswitch phase, the experimenter stated, “Now we are playing a game, an animals game. In this game, all the horses go here (point to box), but the fish go in that box (point).” When the postswitch phase started, children were told, “Okay, now we are going to play a new game, the How-Big game. The How-Big game is different. All the big ones go here (point), but the small ones go in that box (point).”

False belief. A traditional false-belief unexpected-transfer task (Wimmer & Perner, 1983) was administered. Two different picture stories were used: One story involved a girl looking for her chocolate bar that had been unexpectedly transferred. The other story involved a boy looking for his rabbit who had unexpectedly moved to a new location. Half of the children were given a particular story at pretest. The other half received this story at posttest.

First, children had to predict the protagonist’s action based on the protagonist’s false belief. Then, children were shown a picture illustrating the protagonist’s erroneous search in the empty location and were asked to explain this action. The memory of critical story events was checked by four control questions (as in Experiment 1).

Control task. Half of the children received relative-clause tests; the other half received number-conservation tasks (Piaget, 1965). In the relative-clause test, children were shown four pairs of pictures illustrating a girl or a boy who is engaged in a particular activity. The experimenter described one picture; and the children had to describe the other picture. For example, one pair of pictures consisted of a child sitting on a chair and a child sitting on a carpet. The experimenter pointed to one picture and said, “Here is a child who is sitting on a chair.” Then, he pointed to the other picture and said, “And here is a child…” Now, the children had to complete the sentence. There were two sets of pictures. Whether a particular set was given at pretest or at posttest as well as the order of the pictures varied among children. The items were taken from Penner (1999). In the number-conservation task, we used (a) a set of 14 white buttons or (b) a set of 16 green squares \((3 \text{ cm} \times 3 \text{ cm})\). The 14 buttons were arranged in two side-by-side rows of seven objects per row, with the individual buttons in each row in one-to-one correspondence. First, the rows were equal lengths. The children were asked, “In which row are there more buttons? In this one (point) or in that one (point) or are there the same number?” Then, one row was lengthened, the rows were reequalized, and the other row was shortened. Following each transformation resulting in rows of unequal length, the children were asked, “Now, does one row have more buttons than the other, or do they both have the same number of buttons?” In the number-
conservation task involving green squares, two sets of eight squares were arranged into two parallel rows. First, one row was shortened. Then, the same row was lengthened.

Half of the children who had been randomly selected for the number-conservation task received a particular version at pretest. The other half were given this version at posttest. The sequence of the two answer alternatives in the critical question following each transformation varied within one task and between the two task versions.

Training

The training sessions consisted of card-sorting tasks (card-sorting group), stories about false beliefs and false statements (false-belief group), or number-conservation tasks and relative-clauses (control group). In each training group, the experimenter gave positive and negative feedback based on performance and provided explanations. The Appendix provides detailed examples of the training.

Card-sorting group. Each training session consisted of a card-sorting task with three dimension switches and two transfer-sorting tasks. In the card-sorting task, children had to sort first by color, then by number, then again by color, and finally once again by number. Two sets of cards (10 cm × 7 cm) were used. Half of the children received one set of test cards (two yellow apples and one green apple) in the first training session and the other set of test cards (two red houses and one blue house) in the second training session. The other half were given the same sets in reverse order. In each set, the test cards matched each of the two different target cards (one yellow apple and two green apples, one red house and two blue houses) on one dimension. If children failed to sort by the relevant dimension on any trial after one of the three switches in dimension, the experimenter explained how the cards have to be sorted and emphasized the need to redescribe the cards in terms of the opposite dimension.

Subsequently, two transfer-sorting tasks followed. In the first task, although the target cards remained the same (e.g., one yellow apple and two green apples), a new test card was introduced. For example, the children had to sort two yellow flowers first by color and then by number. The experimenter asked the children which dimension was relevant in the color game, for example. Then, the children had to describe the value of the currently relevant dimension displayed on the test card (e.g., “What color are the flowers?”). Finally, the children were required to place the card into one of the two boxes. During the whole procedure positive or negative feedback was given. In the second transfer-sorting task, new target cards (e.g., one blue pear and two red pears) and a new test card (e.g., one red pear) were introduced. Apart from that, the procedure was equivalent to the first transfer-sorting task.

False-belief group. The false-belief training emphasized the fact that mental representations (e.g., false beliefs) and statements (e.g., false utterances) can differ from reality. The false-statement condition was adopted from Hale and Tager-Flusberg (2003) because their training on false statements had a most impressive effect on children’s performance on theory-of-mind tasks. In each training session, two (of four) Ernie-says-something-wrong stories were presented. Three Sesame Street puppets (Ernie, duck, and bear), about 20 cm tall, were used. In each story, Ernie did something to one puppet but said he did it to the other puppet. Children were asked about the content of the statement and about the conflicting reality. The sequence of these two critical questions was varied within children and between children. Half of the children received the four stories in reverse order.

The false-belief story followed Wimmer and Perner’s (1983) traditional unexpected transfer task. One of two stories with Ernie as protagonist was enacted in each session. In Ernie’s absence, his apple (or his shell) was unexpectedly transferred to a new location by the bear. The aim of the training procedure was to highlight those aspects of the unexpected transfer situation that are most significant for understanding. Training should lead the children step by step to an understanding of Ernie’s false belief: It was emphasized that Ernie’s belief resulted from his earlier perception and led to (the wrong) action despite his desire to find it. The direction of transfer (left to right vs. right to left) was counterbalanced across training sessions to avoid causing children to pick up a location-based rule for predicting Ernie’s false-belief-based action.

Control group. The number-conservation training consisted of three parts: familiarization task, spatial transformation, and quantitative transformation. Each training session started with the familiarization task. The order of the spatial transformation and the quantitative transformation varied within children between the two sessions and between children within one training session. In each session, a different set of objects was used (coins or matches). Half of the children started with coins, and half began with matches.
The relative-clause training was a modified version of the relative-clause training used by Hale and Tager-Flusberg (2003). First, Ernie (a rag doll) and two identical-looking twin girl dolls were introduced. In each training session, the experimenter enacted four stories. Each of the twins performed a particular action. Then, Ernie interacted with one of the dolls, and the children had to report with which twin Ernie had interacted. The children received confirming or corrective feedback. Finally, the experimenter reenacted the event and described Ernie’s action using a relative clause.

For example, the following scene was enacted, “This girl is jumping up and down. That girl is shaking her head. Ernie comes. Ernie kisses the girl who was shaking her head.” The twin dolls were hidden under the table, and the children were asked, “Which girl did Ernie kiss?” In the case of no response, the two answer alternatives were provided (“Did Ernie kiss the girl who had been jumping up and down or did Ernie kiss the girl who had been shaking her head?”). The children received positive or negative feedback—“Right” or “No”—“Ernie kissed the girl who had been shaking her head.” Finally, the relevant actions were reenacted: “This girl is shaking her head. Ernie comes and kisses the girl who was shaking her head.” That is, the correct relative clause was modeled once again.

In the second session, in each story the actions of the twins remained the same, but now, Ernie performed another action, and he interacted with the other twin. Half of the children received the eight stories in reverse order.

Posttest

A card-sorting task, a false-belief task, and a control task were administered. In addition, a three-boxes card-sorting task was used. The procedure was similar to the standard DCCS. However, 3 target cards (a red horse, a blue fish, and a green bird) and 21 test cards (7 blue horses, 7 green fish, and 7 red birds) were employed. Children were required to sort 9 cards in the preswitch phase and postswitch phase. Each target card was affixed on a (22 cm × 15 cm × 14.5 cm) box. Half of the children sorted first by color and then by shape. The other half began with sorting by shape. The preswitch phase and the postswitch phase matched the standard procedure except that in each sorting phase three sorting boxes (instead of two) were used. Use of three boxes makes it possible to distinguish whether a training effect on the DCCS task is due to applying a reversal shift on the preswitch dimension because such a reversal shift strategy cannot be applied to a card-sorting task with three boxes.

Half of the children were first given the false-belief task and then given the three-boxes card-sorting task, the control task, and the card-sorting task. The other half received the reverse order.

Results

Scoring

On each DCCS task, children were given a score between 0 and 5 depending on the number of cards sorted correctly in the postswitch phase. On the control tasks, a score between 0 and 4 could be obtained: One point was given for each correctly formed relative clause (out of four), and 2 points were given for each correct answer to one (out of two) test questions in the number-conservation task.

On each false-belief task, children received a score of 0, 1, or 2 based on the prediction and explanation test questions. Answers to the open explanation question were classified in the following categories: (a) mental state, for example, “He thought it was in here,” “He doesn’t know it’s in there,” “He didn’t see it being moved”; (b) relevant story facts, for example, “He had been away,” “It was in here earlier”; (c) desire answers, for example, “because he wants it”; (d) wrong-location answers, for example, “because it isn’t here,” “because it is over there”; (e) moved-away answers, for example, “because he has moved it away,” “because the rabbit moved away”; (f) no answer or “don’t know” answer. For further analysis, answers of categories (a) and (b) were classified as correct answers indicating an understanding of belief. The answers in the remaining categories were classified as incorrect (Wimmer & Mayringer, 1998).

On the three-boxes card-sorting task, a score between 0 and 9 could be obtained reflecting the number of correct sorts in the postswitch phase. All scores were transformed into values between 0 and 1 indicating the proportion of correct answers.

Pretest Performance

Table 2 presents the pretest scores of the children in each of the training groups. One-way ANOVAs revealed no significant differences on any of these and other relevant variables: age, $F(2, 41) = .34$, $p > .70$; verbal intelligence, $F(2, 41) = 2.12$, $p > .10$; card-sorting pretest score, $F(2, 41) = .12$, $p > .80$; false-
belief pretest score, \( F(2, 41) = .19, p > .80 \); control task pretest score, \( F(2, 41) = .13, p > .80 \).

**Training Effects**

Patterns of performance at pretest and posttest along with the observed changes are shown in Table 2.

**Card Sorting**

First, a repeated measures ANOVA was conducted on the card-sorting performance with pre-post (pretest score, posttest score) as a within-participants factor and training method (card sorting, false belief, and control training) as a between-participants factor. There was a significant main effect of pre-post, \( F(1, 41) = 53.84, p < .001 \), and a significant Pre-Post \( \times \) Training Method interaction, \( F(2, 41) = 3.55, p < .05 \). The significant main effect of pre-post was further analyzed by one-tailed paired \( t \) tests (with Bonferroni correction of significance levels for post hoc tests). The pre-post comparison was significant in the card-sorting group, \( t(13) = 5.26, p < .01 \), and in the false-belief group, \( t(14) = 3.07, p < .05 \), but not in the control group, \( t(14) = 1.99, p > .05 \). A one-way ANOVA with pre-post difference as the dependent variable revealed that the significant interaction between pre-post and training method was due to the fact that the improvement in the card-sorting group was significantly greater than in the control group (least significance difference [LSD] test: \( p = .011 \)). Improvement in the false-belief group did not differ significantly from improvement in either of the other two groups.

On the three-boxes card-sorting task, children in the card-sorting group (93% correct) performed better than the other two groups (false-belief group: 70% correct; control group: 64% correct). A one-way ANOVA with the three-boxes card-sorting task as the dependent variable revealed an almost significant effect of training group, \( F(2, 41) = 3.01, p = .06 \). LSD tests showed only a significant difference between the card-sorting group and the control group (\( p < .05 \)).

**False Belief**

A repeated measures ANOVA was also conducted on the false-belief scores, showing only a significant main effect of pre-post, \( F(1, 41) = 11.28, p < .005 \). The Pre-Post \( \times \) Training Method interaction was not significant, \( F(2, 41) = .22, p = .80 \). Bonferroni corrected one-tailed paired \( t \) tests indicated that the card-sorting group showed a significant improvement on the false-belief task, \( t(13) = 2.46, p < .05 \), but not the false-belief group (\( p = .13 \)) or control group (\( p = .18 \)).

**Control Tasks**

A repeated measures ANOVA revealed no significant effects for the control task in any of the training groups. Separate pre-post comparisons also failed to show any pre-post improvements for card-sorting or false-belief training. However, Bonferroni corrected paired \( t \) tests showed that control training did produce significant improvement on relative clauses from pretest (\( M = 57% \) correct) to posttest (\( M = 82% \) correct), \( t(6) = 4.58, p < .01 \), one-tailed. But children in the number-conservation subgroup (\( n = 8 \)) showed no significant improvement on number conservation from pretest (\( M = 19% \) correct) to posttest (\( M = 31% \) correct). In the false-belief and card-sorting groups, separate analyses revealed no significant change in performance on the relative-clause or number-conservation task.

**Other Effects**

Replicating previous studies (Clements et al., 2000; Perner, Lang, et al., 2002), there was a

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Percent Correct on Pretest, Posttest, and Their Difference (Δ) on the Three Tasks for Each Training Group of Experiment 2</td>
</tr>
<tr>
<td><strong>Training group</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Card sort</td>
</tr>
<tr>
<td>False belief</td>
</tr>
<tr>
<td>Control</td>
</tr>
</tbody>
</table>

*Note. Δ = posttest–pretest difference; DCCS = Dimensional Change Card Sorting task.*

*p < .05. **p < .01.
significant gap between children’s ability to predict false-belief-based action and their ability to explain the erroneous action at pretest. McNemar’s $\chi^2(1, N = 44) = 16.96, p < .001$, and at posttest, McNemar’s $\chi^2(1, N = 44) = 25.04, p < .001$. There were only two children at pretest (and no child at posttest) who were able to answer the explanation question while failing the prediction question. However, 24 children at pretest and 27 children at posttest failed to give satisfactory explanations despite correct predictions.

In line with this, 61% of the children (67% in the false-belief group, 57% in the card-sorting group, and 60% in the control group) were able to predict the protagonist’s false-belief-based action at pretest. At posttest, a large majority of children in all three training groups answered the prediction question correctly (false-belief group: 87%; card-sorting group: 93%; control group: 80%). In contrast, only 5 children were able to explain the protagonist’s erroneous action at pretest. Three children referred to a mental state, and 2 children mentioned relevant story facts. The incorrect answers consisted of 16 desire answers, 10 wrong-location answers, 2 moved-away answers, and 11 “no” or “don’t know” answers. There was only minimal improvement in explanations between pretest and posttest: Only 3 children in each training group were able to answer the explanation question at posttest after having given unsatisfactory explanations at pretest. Moreover, 1 child in each training group failed to give a satisfactory explanation at posttest even after having done so at pretest. In sum, 4 children gave mental state answers, and 7 children made note of relevant story facts. Incorrect answers consisted of 13 desire answers, 11 wrong-location answers, 2 moved-away answers, and 7 “no” or “don’t know” answers.

**Discussion**

This experiment showed that a card-sorting training with feedback and explanations markedly improves children’s performance on the DCCS task compared with a control group. Furthermore, a false-belief training also led to a significant rise in card-sorting performance, though not more so than in the control group. In addition, the card-sorting training significantly increased children’s performance on the false-belief task, although not significantly more than control training. The mutual transfer effects between the false-belief task and the DCCS task provide the first evidence that there is a causal link between theory of mind and executive function (at least as exemplified by these two tasks) that was previously only indicated in correlational studies (e.g., Carlson & Moses, 2001; Frye et al., 1995; Perner, Lang, et al., 2002).

One surprising outcome of the present study is that false-belief training did not significantly enhance false-belief performance though it improved card-sorting performance. An explanation for this finding may be children’s good performance on the prediction question at pretest, leaving little room for improvement on this question. Furthermore, correct explanations might have been out of the children’s “zone of proximal development” (Clements et al., 2000; Vygotsky, 1978).

The transfer effect on the three-boxes card-sorting task indicates that children learned something that goes beyond the immediate context of the two-boxes task. The card-sorting training with two boxes might have led children to entertain a simple strategy about the switch in rules: “If the experimenter suggests playing a new game, I have to reverse my actions.” But this strategy cannot be applied to the three-boxes card-sorting task. Yet, children in the card-sorting group were almost perfect (93% correct) on this task and performed significantly better than the control group.

The significant difference in improvement between the card-sorting group and the control group indicates that the training effect of the card-sorting training cannot be attributed to maturation or contact with the experimenter. However, the exact nature of the training effect remains to be determined. We do not know for sure that redescription is at the heart of the training effects. To show this, a more stringent control condition would be needed, for example, giving children the DCCS task without explicit training or feedback.

**GENERAL DISCUSSION**

The main finding of the reported experiments is that children’s difficulties with the standard DCCS task can be effectively remedied by giving them feedback and explanations about the need to redescribe the cards to be sorted. This training also had some effect on their ability to pass the false-belief test. Moreover, training on the false-belief problem had some beneficial effect on their ability to sort cards correctly. From this the following questions arise: What is children’s problem with the card-sorting task and the false-belief task that this training helps overcome? Is it a fairly specific problem shared by these two tasks or is it a much more general problem that pertains to a wide range of executive-function and theory-of-mind tasks?
Recent evidence on the DCCS task points to a specific problem, namely, a difficulty with redescribing the cards to be sorted (e.g., Experiment 1 reported here; Kirkham et al., in press; Perner & Lang, 2002; Towse et al., 2000). There is independent evidence that young children before the age of about 4 years, before they understand false belief, have difficulty understanding that objects can be redescribed. Although children fairly early acquire different names for things (Clark, 1997), they are reluctant to use both terms at the same time, as reflected in the well-known mutual exclusivity bias (Carey & Bartlett, 1978; Markman & Wachtel, 1988). Doherty and Perner (1998) asked children to acknowledge explicitly that something can be two things at the same time, for example, that something can be a rabbit and a bunny, or a rabbit and an animal (Perner, Stummer, Sprung, & Doherty, 2002) and found that this ability emerges with their ability to understand false belief.

Similarly, Flavell (1988) characterized 3-year-olds' problems in theory-of-mind tasks in the following way:

They tend to be largely ignorant of the fact that it is possible to represent a single thing with its single nature in several different ways—ways that would be mutually contradictory if they described the object itself rather than mental representations of it. (p. 246)

Finally, children’s inability to understand redescription, we suggest, is the common denominator underlying children's difficulty with the standard DCCS task and their failure to understand false belief. To understand false belief, one has to understand that someone else can have a description of the real world that differs from one’s own description. This explains why performance on the DCCS task tends to be strongly correlated with mastery of the false-belief task (e.g., Carlson & Moses, 2001; Frye et al., 1995; Perner, Lang, et al., 2002) and why the card-sorting training in our Experiment 2 tended to have an effect on children’s understanding of belief and vice versa.

If we are correct and the correlation between performance on false-belief and card-sorting tasks is due to the ability to redescribe objects as different things, that raises the question of why there is also a correlation between false-belief tasks and other executive-function tasks that do not require such redescription (e.g., Carlson & Moses, 2001; Hughes, 1998). Most of these tasks, for example, Luria’s hand game or the bear–dragon task (Reed, Pien, & Rothbart, 1984), seem to involve no need to describe one and the same entity differently. Therefore, the common denominator between theory of mind and executive functions in general cannot be the ability to describe one and the same thing differently. This raises the following question: Are these correlations due to other specific factors shared by the false-belief task and these other executive tasks, or can we find a more general common denominator for all these tasks? As noted earlier, several theories have been proposed to explain the relation between theory of mind and executive function in general. The present data give some help in deciding among them.

The explanation that theory-of-mind tasks correlate with executive tasks because they contain executive components was weakened by the fact that the explanation version of the false-belief task correlates with executive tasks as strongly as the prediction version (Hughes, 1998; Perner, Lang, et al., 2002). Furthermore, the explanation version is more difficult than the prediction version (e.g., Experiment 2; Clements et al., 2000). This indicates that response inhibition cannot be the main or only problem for children in the false-belief task because there is no prepotent response tendency when asked to explain a false-belief-based action.

The explanation that maturation of common brain regions accounts for the correlation between false belief and executive functions was ruled out by our training effects. Only if we allow for the possibility that our training stimulated brain development will this explanation stay viable. This would be particularly plausible for a relation between theory-of-mind tasks and Stroop-like executive tasks requiring inhibitory processes (Carlson & Moses, 2001), as the physiological basis for the skills needed in these tasks has been identified in neighboring regions in ACC. However, the evidence on what makes the DCCS task difficult does not support a Stroop-like interference of two action tendencies (as discussed previously).

Zelazo and Frye (1997; CCC theory) proposed that the ability to reason with embedded conditionals may be demanded both by theory-of-mind tasks and by rule-use tasks, such as the DCCS. However, several experiments (e.g., Experiment 1; Perner & Lang, 2002; Towse et al., 2000) have indicated that, strictly speaking, children’s difficulty in the DCCS cannot be explained by an inability to employ an embedded if-if-then rule structure. CCC theory might be able to explain these results with the additional stipulation that the use of a higher order rule becomes problematic for children only if it
involves a problem of identity as alluded to in the following passage by Zelazo and Frye (1997):

Higher-order rules allow children to understand that the shape rules and the color rules both apply to a single task under different descriptions (or setting conditions). Likewise, the difficulty that children have distinguishing between appearance and reality is overcome when children condition- 

alize “identities” on setting conditions and integrate the identities into a single system of ideas, much as the morning star and the evening star come to be connected to one’s ideas about Venus. (p. 145)

Still, a problem remains when we realize that the identity of objects poses the critical difficulty for younger children. This is explicitly shown for the card-sorting task in the study by Kloo and Perner (2002). The question for the amended CCC theory is which object needs to be redescribed in the ramp task (Frye et al., 1995), in which the configuration of a ramp apparatus determines whether a marble will roll straight down or crosses over. CCC theory needs to give an answer to this question because children’s difficulties on the ramp task have been seen as core evidence for the theory from its inception.

The reciprocal transfer observed in Experiment 2 is compatible with both theories, which see a functional dependency between the development of theory of mind and executive functions. Additional lines of evidence will be needed to differentiate between these two theories. Especially, longitudinal studies and training studies employing different executive tasks and various theory-of-mind tasks may provide further evidence. However, it may be that both functional-dependence theories are right in the sense that theory of mind and executive functions are interdependent in their development.

In sum, we can tentatively offer the following options for explaining the developmental link between theory of mind and executive development. First, specific factors underlie particular tasks, for example, understanding of redescription of objects and events may underlie the card-sorting and false-belief tasks. Other specific factors need to be identified for explaining correlations reported for other executive-function and theory-of-mind tasks. Second, there is a general functional link. Understanding the mind presupposes a certain degree of executive control, and executive functioning presupposes a certain level of insight into the mind. This could explain the correlational data and the results of our training study, but it sheds no light on the reasons and processes that create these presuppositional relations.

References
Appendix

Card-Sorting Training

If children failed to sort by the relevant dimension (e.g., they failed to sort a card with two red houses
on it by number) on any trial after one of the three switches in dimension, the experimenter stated, “No, that was wrong. You put the red one into the box with red on it, that is, you looked at the color. However, we are not playing the color game, the game with red and blue (point), anymore. Now, we are playing the How-Many game. This is the game with one and two (point).”

Question (Q): What game are we playing now?
Feedback (F): Right/No. We are playing the How-Many game now. This is the game with one and two (point). Then, the experimenter emphasized the need to redescibe the cards in terms of the opposite dimension, “So, you have to look how many houses there are on the card,” and asked the children to label a test card (with two red houses on it) by the new dimension.

Q: Look, how many are on this card?
F: Right/No. There are two.
Then, children were asked to label the corresponding target card.

Q: Look, how many are on this card?
F: Right/No. There are two.
The experimenter repeated the relevant one of the two rules (“In the How-Many game, all the cards with two on it go into the box with two on it”), placed the test card into the appropriate box and asked the children about that rule.

Q: Where does the card with two on it go in the How-Many game?”
F: Right/No. The card with two on it goes into the box with two on it.
Finally, children were handed the corresponding target card.

Q: There are two. Where does this card go in the How-Many game?

False-Belief Training Group

Ernie-Says-Something-Wrong Story

Ernie kisses the duck and says to the child, “I kissed the bear!”

Q: What did Ernie really do? (Prompt, if necessary: Who did Ernie really kiss?)
F: Right/No. Ernie kissed the duck.
Q: What did Ernie say? (Prompt, if necessary: Who did Ernie say he kissed?)
F: Right/No. Ernie said that he kissed the bear.
The experimenter reenacted the story and said, “Ernie kissed the duck. Then, he said, ‘I kissed the bear.’ But really, he kissed the duck.”

False-Belief Story

Ernie comes. He has got a shell. Ernie puts the shell in the yellow tower.

Prompt: Where did Ernie put the shell?
Then, the duck comes and asks Ernie, “Ernie, where is your shell?”
Q: What will Ernie say to the duck?
F: Right/No. Ernie says, “The shell is in the yellow tower.”
Q: Where does Ernie think the shell is?
F: Right/No. Ernie thinks that the shell is in the yellow tower because he put it there.

Then, Ernie goes together with the duck to the swimming pool. While Ernie and the duck are swimming, the bear comes. He takes the shell out of the yellow tower and puts it in the red house. Then, the bear leaves.

Q: Where is the shell now?, Who put it there?, Was Ernie able to see this?
Look, there are Ernie and the duck. They are talking about the shell. The duck asks Ernie, “Ernie, where did you put the shell earlier on?”
Q: What will Ernie say to the duck?
F: Right/No. Ernie says, “I put the shell in the yellow tower.”
Q: Where did the bear put the shell?
F: Right/No. He put it in the red house.
Q: Could Ernie see that?
F: Right/No. Ernie did not see that.
Q: So, does Ernie really know that the shell is in the red house?
F: Right/No. Ernie does not know that the shell is in the red house now.

Q: Where does Ernie think the shell is?
F: Right/No. Ernie still thinks that the shell is in the yellow tower.

The duck says to Ernie, “Let’s get the shell.”
Q: Where will Ernie go first?
F: Right/No (look). Ernie goes to the yellow tower.

Q: Why does Ernie go to the yellow tower when he wants to get the shell?
F: Right/No. Ernie goes to the yellow tower because he still thinks that the shell is in the yellow tower.

Q: Why does Ernie still think that the shell is in the yellow tower?
F: Right/No. Ernie did not see that the bear put the shell in the red house. So, he does not know that the shell is in the red house now. Ernie looks into the yellow tower, “Oh! The shell is not there. But I put it there. I thought that the shell is in the yellow tower. Where is the shell?! I do not know at all where it is.”
Q: What should Ernie do?

Control Training—Number-Conservation Training

Familiarization Task

Two rows of white plastic pearls were laid out. The shorter row consisted of six pearls. The longer row contained five pearls. The children were asked to choose a row, “Look, there are two rows of pearls. One of the two rows is for you!” Reliance on the length of the rows alone would result in choosing the row containing only five pearls. No feedback was given. The arrangement of the two rows (i.e., which row was closer to the child) varied within and between children.

Spatial-Transformation Task

The procedure was similar to the number-conservation task used at pretest and posttest. For example, two parallel rows, each containing five coins, were laid out.

Q: In which row are there more coins? In this one (point) or in that one (point) or are there the same number? When the child agreed that the number of coins was equal, one row was shortened.

Q: Now, does one row have more coins than the other, or do they both have the same number of coins?

F: Right/No. Both rows have the same number of coins because I did not take any coin away, and I did not add one. I just shortened this row (point). The rows were equalized again, and the experimenter said, “Look! Both rows have the same number of coins.”

Quantitative-Transformation Task

Children were still presented with the two parallel rows, each containing five coins.

Q: In which row are there more coins? In this one (point) or in that one (point) or are there the same number? When the child agreed that the number of coins was equal, a coin was subtracted from one of the rows, by keeping the length of the rows the same.

Q: Now, do they both have the same number of coins, or does one row have more coins than the other?

F: Right/No. This row (point) has more coins because I took away a coin from the other row (point). Then, the individual coins in each row were arranged in one-to-one correspondence resulting in one coin standing alone. The experimenter said, “Look! There is one more coin.” Then, the subtracted coin was added again, and the experimenter said, “Look! Now, both rows have the same number of coins.” Then, a coin was added to the other row, by keeping the length of the rows the same.

Q: Now, does one row have more coins than the other, or do they both have the same number of coins?

F: Right/No. This row (point) has more coins because I added one coin. Again, the individual coins in each row were arranged in one-to-one correspondence and the experimenter said, “Look! There is one more coin.” Then, the added coin was subtracted again, and the experimenter stated, “Look! Now, both rows have the same number of coins.”

The row that was manipulated, the order of the two possible transformations, and the sequence of the two answer alternatives in the critical questions varied between training sessions in both the spatial and the quantitative-transformation task.