Young Children's Differentiation of Hypothetical Beliefs from Evidence

Author(s): Beate Sodian, Deborah Zaitchik and Susan Carey

Source: *Child Development*, Vol. 62, No. 4 (Aug., 1991), pp. 753-766
Published by: Wiley on behalf of the Society for Research in Child Development
Stable URL: http://www.jstor.org/stable/1131175
Accessed: 28-08-2017 18:47 UTC

Young Children's Differentiation of Hypothetical Beliefs from Evidence

Beate Sodian

University of Munich and Massachusetts Institute of Technology

Deborah Zaitchik and Susan Carey

Massachusetts Institute of Technology

SODIAN, BEATE; ZAITCHIK, DEBORAH; and CAREY, SUSAN. *Young Children's Differentiation of Hypothetical Beliefs from Evidence*. CHILD DEVELOPMENT, 1991, **62**, 753–766. The claim that preadolescent children fail to differentiate between hypothetical beliefs and evidence is investigated in 2 studies. First- and second-grade children were presented with 2 conflicting hypotheses and asked to choose an empirical test to decide between them. In Study 1, the majority of first graders and almost all second graders correctly chose a conclusive test. They elaborated the logic of such a test and distinguished it from an inconclusive test. There was no evidence that children of this age misinterpret the task of hypothesis testing as one of generating a desirable effect. Study 2 replicated and extended these findings; in a task that posed a genuine scientific problem, first and second graders spontaneously generated empirical procedures for gathering indirect evidence to decide between alternative hypotheses. Our results indicate that young elementary school children distinguish between the notions of "hypothetical belief" and "evidence." These findings are discussed in light of their failure on other scientific thinking tasks.

Elementary school children are generally considered to lack an understanding of the "scientific method." That is, while young children may construct intuitive theories of the world, they lack metaconceptual awareness of this fact and have little understanding of the components of scientific reasoning (e.g., Dunbar & Klahr, 1989; Inhelder & Piaget, 1958; Kuhn, 1989; Kuhn, Amsel, & O'Loughlin, 1988; Kuhn & Phelps, 1982). Asked to determine the causes of a particular phenomenon, they often fail to test hypotheses in a systematic way; instead, they act as if their goal were simply to produce or repeat the effect, rather than to discover its causes. When evaluating hypotheses concerning the effects of particular variables, they often fail to control for confounding variables. When interpreting the results of their "experiments," they usually account for only parts of the available data, neglecting evidence that conflicts with their currently favored interpretation. Furthermore, they hardly ever explicitly seek such disconfirming evidence.

One interpretation of such findings is that young children have limited metaconceptual understanding of notions like "hypothesis" and "test." Deanna Kuhn and her colleagues (Kuhn, 1989; Kuhn et al., 1988) characterized the child's metaconceptual understanding of the nature of science as severely deficient, lacking the fundamental differentiation between the notions of "theory" and "evidence." According to Kuhn, preadolescent children are unable to distinguish between their beliefs and evidence supporting or disconfirming their beliefs. Instead of seeing evidence that supports their belief as providing confirmation, they seem to view it as an *instance* of the state of affairs which their belief proposes. When presented with contradictory evidence, they typically either distort the evidence to adjust it to their prior beliefs or they adjust their beliefs without being aware of doing so. In recent work, Kuhn has found that even some adults, when asked for evidence, merely elaborate their theories. For example, asked for evidence for the belief

Beate Sodian acknowledges support from the Deutsche Forschungsgemeinschaft (So 213/5-1). Deborah Zaitchik and Susan Carey acknowledge support from the McDonnell Foundation. We are grateful to the teachers and children of Peabody School, Cambridge, and Solomon Schechter Day School, Newton, for their friendly cooperation. Requests for reprints should be addressed to Susan Carey, Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139 or to Beate Sodian, University of Munich, Leopoldstr. 13, 8000 München 40, Germany.

[*Child Development*, 1991, **62**, 753–766. 0009-3920/91/6204-0015$01.00]

that recidivism by criminals is due to their inability to cope in the real world, a subject might elaborate the ways in which criminals cannot cope. Based on such findings, Kuhn (1989) suggests that children do not distinguish between their own theoretical ideas and external sources of evidence relevant to these ideas, and consequently do not understand what it means to "test" an idea or to "prove" a claim (see also Dunbar & Klahr, 1989; Dunbar et al., 1989).

The present studies explore Kuhn's claim. To assess whether the child distinguishes the notions of "theory,[1] belief, hypothesis," on the one hand, from the notion of "evidence" on the other, we must begin with an analysis of these notions. Beliefs are mental representations of reality, for example, the beliefs that the earth is a sphere or that eating cabbage protects against cancer. Hypotheses are beliefs that are subject to confirmation or disconfirmation. To test a hypothesis means to generate evidence relevant to its confirmation or disconfirmation. Evidence, for example, a picture of the earth from a spaceship or data from a study showing a lower frequency of cancer in people who eat cabbage than in those who abhor cabbage, is inferentially related to hypothetical beliefs. To frame a belief as a hypothesis requires imaging alternative states of the world and judging whether each is consistent with the hypothesis or not. The goal of hypothesis testing is to generate evidence that conclusively confirms or disconfirms the hypothesis tested, that is, evidence that allows one to decide between alternative hypotheses. To meet this goal, the hypothesis tester has to specify how alternative states of the world would bear on the truth or falsity of the hypothesis in question, that is, which inferences would be warranted on the basis of specific evidence. Clear support for the position that children possess a metaconceptual understanding of the notions of "hypothesis" and "evidence" would come from a demonstration of their ability to explicitly lay out the inferential relation between hypotheses and evidence. To do this, children need to understand that evidence is the basis from which inferences are drawn about the truth or falsity of a hypothesis, and they must know that the criterion for evaluating evidence is whether it supports, undermines, or is irrelevant to a hypothesis (or to the choice between alternative hypotheses).

In their second year of life, children begin to exhibit behavior that Piaget interpreted as "experimenting." For example, they let an object fall on the floor repeatedly, "systematically" varying the height and observing the effect. These infants may entertain "hypotheses," for example, about the relation between the height at which they let the rattle fall and the noise it makes when it hits the floor. Furthermore, 18-month-olds *use* evidence, even indirect evidence, in their logical search behaviors (Haake & Somerville, 1985). However, these data do not warrant attributing a metaconceptual understanding of the notions of "hypothesis," "test," and "evidence" to such young children. Specifically, Sophian and Somerville (1988) demonstrate that it is not until age 4 that children maintain alternative hypotheses concerning the location of a hidden object. And even these demonstrations do not show that 4-year-olds understand how hypothetical states of affairs would bear on the truth of such hypotheses. To understand this the 4-year-old would have to understand which inferences are warranted on the basis of the available evidence. The theory of mind literature shows that by the age of 4, children understand that knowledge and belief are based on perceptual experience; they distinguish beliefs and reality and consequently appreciate that beliefs may be true (correspond to reality) or false (fail to correspond to reality) (see Perner, in press; Wellman, 1990, for reviews). However, they do not understand *inference* as a source of knowledge until age 6. Four-year-olds attribute ignorance to a person who has not witnessed a hiding event but has sufficient information to infer its outcome (Sodian & Wimmer, 1987). Thus, although 4-year-olds proficiently draw inferences based on the evidence available to them, they do not seem to understand that inferences are warranted on the basis of available evidence.

If preschoolers do not understand that knowledge can be acquired inferentially, one would not expect them to distinguish between warranted and unwarranted inference. Thus, it is not surprising that although preschoolers show some sensitivity to inde-

[1] The studies of Kuhn et al. (1988) investigate an understanding of hypotheses about causal relations and covariation evidence relevant to these hypotheses, not a metaconceptual understanding of *theories* as sets of interrelated beliefs. Similarly, we do not attempt to explore children's grasp of the notion of a "theory" (as opposed to the notion of a simple belief).

terminacy in their own search behaviors (Fabricius, Sophian, & Wellman, 1987; Sophian & Somerville, 1988), they do poorly on tasks that require them to recognize the effects of indeterminacy on a searcher's knowledge. Somerville, Hadkinson, and Greenberg (1979) told children, for instance, that "the boy lives in the house with the blue table in the front yard" when there were in fact *two* houses with blue tables in the front yard. The 6-year-olds, but not the 5-year-olds, could judge that they needed to ask for further information before they could specify the correct location. Similarly, studies of children's ability to recognize referential ambiguities show that although preschoolers themselves are uncertain about the correct referent when hearing ambiguous messages, only 6-year-olds correctly distinguish between the effects of ambiguous and unambiguous messages on some other listener's knowledge (Robinson & Whittaker, 1985; Sodian, 1988).

Thus, the theory of mind and the logical reasoning literatures[2] indicate that by the age of about 6 years children possess the conceptual prerequisites for understanding how evidence bears on hypotheses: They understand that information about some state of the world can give rise to inferences about other states of the world, and they can distinguish between informational conditions that allow one to draw a certain inference and those that do not. However, it has not been shown that children understand that to test a hypothesis means to generate evidence that confirms or disconfirms it. Such a task, requiring children to devise and evaluate tests of hypotheses (e.g., requiring them to distinguish between conclusive and inconclusive tests), would provide an adequate assessment of children's ability to differentiate between the notions of "hypothesis" and "evidence."

Karmiloff-Smith and Inhelder (1974/75) showed that young elementary school children distinguish in their action between the goals of producing an effect (making a block balance) and of testing a theory (of why a given block balances where it does). Karmiloff-Smith and Inhelder endorse the view that these young children had no metaconceptual awareness of this distinction, since "the younger child clearly cannot reflect on hypothetical situations which might confirm or refute his theory" (p. 209). However, metaconceptual understanding of hypothesis testing was not explicitly probed for in this study. It was the focus of a study by Tschirgi (1980), who presented children with a choice between three strategies (confirming, disconfirming, confounding) to "prove" that a particular variable (e.g., flour) was the cause of either a positive or a negative outcome of some everyday activity (e.g., baking a cake). When testing hypotheses about the causes of negative outcomes (the bad cake), 9–10-year-olds clearly preferred a disconfirming strategy (vary only the hypothesized variable, e.g., the flour, and hold all others constant). Even second graders showed some limited competence in this task. However, when asked to test a hypothesis about a positive outcome (a good cake), both children and adults tended to choose a confirming strategy (hold hypothesized variable, the flour, constant and vary all others). These data might be taken as support for Kuhn's claim, since they are consistent with the lack of a metaconceptual distinction between testing a hypothesis and producing an effect. After all, children's choice of a sensible strategy in the negative outcome condition may only reflect their ability to avoid a bad outcome by changing the variable they take to be causal while ignoring those they do not take to be causal. Indeed, adults' choice of a confirming strategy in the positive outcome condition supports Kuhn's contention that even some adults do not have a firm concept of evidence as related to experimental tests. This is, however, not the only interpretation possible. Tschirgi's findings can also be taken to indicate that elementary school children understand in principle the goal and logic of testing a hypothesis,[3] but that a strong pragmatic bias to "keep a good

[2] In the literature on the development of logical reasoning, it has been claimed that only around the age of about 10 years do children become able to distinguish between necessity and possibility and hence to distinguish between warranted and unwarranted inference (e.g., Pieraut-LeBonniec, 1980). However, Braine and Rumain (1983) conclude from a review of the literature that children's failures on tasks requiring this distinction seem to be attributable to a response bias of avoiding a "can't tell" response in tasks that lead them to expect a single solution.

[3] Similarly, studies by Case (1974), Siegler and Liebert (1975), and Siegler, Liebert, and Liebert (1973) showed that a systematic control-of-variables scheme can be acquired by children around the age of 10 years (or even by gifted 8-year-olds) after some degree of instruction and practice.

result" masks both children's and adults' competence in the positive outcome condition. In a task similar to Tschirgi's, Bullock (1989) asked children to justify their strategy choice. While the justifications of the 7-year-olds referred mostly to the goal of producing a good result, those of the 8- and 9-year-olds referred to the task of testing a hypothesis. It seems, then, that 8-year-olds use a disconfirming strategy in order to generate conclusive evidence about the hypothesis rather than simply to produce a desirable effect. Still, there were many result-oriented responses from the older children. It is possible, then, that Kuhn is correct in claiming that young elementary school children do not differentiate beliefs from evidence. In other words, despite their command of the relevant prerequisites, they may not yet have acquired the concepts "hypothesis testing" and "evidence."

The goal of the present studies is to test whether young elementary school children are, in fact, fundamentally unable to distinguish between hypotheses and evidence. We do so by posing the following two questions: (1) Do children have the notion of testing a hypothesis as opposed to producing an effect? and (2) Given a choice between conflicting hypotheses, can they distinguish between experiments that would produce conclusive as opposed to inconclusive evidence? To answer these questions, we designed tasks that tap the relevant distinctions while avoiding many of the additional demands often found in scientific thinking tasks. For instance, children are typically presented with hypotheses and evidence about *causal* relations among variables. In such cases, evidence is consistent or inconsistent with hypotheses, but the outcome of a single test can never be conclusive. Furthermore, the child must understand the isolation of variables, the importance of controlling for confounds, and the necessity of seeking counterexamples to master such tasks. Also, in domains where children have favored hypotheses of their own, they must generate, or at least recognize, alternative hypotheses. While we agree with other investigators that all these abilities are important components of scientific reasoning, they are not necessary for making a fundamental distinction between hypotheses and evidence. To avoid these additional demands, we devised tasks in which the hypotheses to be tested are about determinate states of affairs (e.g., the size of an object); in our tasks, then, a single test can provide conclusive evidence. Furthermore, the hypotheses we chose made it unlikely that children should have any previous bias toward any of them. We carried out two studies with first- and second-grade children. To assess within-child consistency, we used the same group of children in both studies; the studies were presented in counterbalanced order in a single session. Study 1 was designed to probe for children's ability to distinguish between testing a hypothesis and producing an effect; here the task of choosing an adequate test for a hypothesis required children to distinguish between a conclusive and an inconclusive test. While the task used in Study 1 posed a contrived situation, Study 2 explored children's ability to spontaneously generate, and/or to choose from two alternatives, an adequate test for a genuine scientific hypothesis.

Study 1

In Study 1, the child was told a story about two brothers who knew they had a mouse in the house (which they had not been able to observe), but who differed in their beliefs as to the size of the mouse. Two boxes were described, one with a big opening and one with a small opening. The child was asked which box the brothers should put out if they wanted to tell, according to whether the food was gone in the morning, which brother was correct in his assumption about the size of the mouse. One possible reason for failure on this task is that children interpret the task of testing hypotheses about the size of the mouse as a task of feeding the mouse (i.e., of producing some effect). To test whether children can distinguish between these two goals, we gave them a second task, in which they were asked to choose which box they should put out if their goal were to make sure that the mouse got some food, no matter the size of the mouse.

An orientation toward producing an effect is not the only possible reason for failure on the hypothesis-testing task. Children may understand the purpose of hypothesis testing but still choose the wrong box because of a failure to recognize that the outcome of a test using the box with a big hole is inconclusive. Note that the task of choosing a conclusive test of a hypothesis requires children to anticipate the epistemic consequences of the two proposed tests (i.e., recognition of the inconclusiveness of one test and understanding of the inferences that can be drawn from either a positive or a negative outcome of the other test). Although 6-year-olds have been shown to recognize indeterminacy, they may not be able to take indeterminacy into account spontaneously when asked to test a hypothesis. Thus, their spontaneous choice of a suitable method for testing the hypotheses about the size of the mouse is a conservative measure of their understanding of the task. They may, however, understand the consequences of indeterminacy for hypothesis testing, once they have worked through the logic of the task. Therefore, for children who initially chose an inconclusive test but then correctly elaborated what follows from a conclusive and an inconclusive test, the initial question was repeated.

Method

Subjects

Subjects included 34 children, 15 boys and 19 girls from two elementary schools in the Boston area. Twenty first graders (M = 6-10, range 6-1 to 7-11) and 14 second graders (M = 8-3, range 7-8 to 9-5) participated in the experiment.

Procedure and Design

The experimenter first explained that she would read a story about two boys who wanted to do an experiment. Each child was then presented with the following story, accompanied by illustrations. "This is Jim and this is Steven. They're brothers. They know there is a mouse in their house, but they never see it because it only comes out at night. They know there is just one mouse. Jim thinks it's probably a big daddy mouse, but Steven thinks it's probably a little baby mouse."

After this introduction, children were presented with the Feed and Find Out conditions, in counterbalanced order:

Feed condition: "The boys decide they want to feed the mouse. Here are two boxes. One has a large opening and one has a small opening. The boxes aren't traps; they're just plain old boxes, like mouse houses. If the mouse can fit through the opening, he can go inside the house, eat the food, and then leave again."

Control questions: "Can a big mouse fit in the house with the big opening? Can a small mouse fit in the house with the big opening? Can a big mouse fit in the house with the small opening? Can a small mouse fit in the house with the small opening?"

"The boys want to leave food for the mouse in one of the boxes. Remember, they don't know whether the mouse is a big daddy mouse or a small baby mouse."

Feed question: "Which box should they use if they want to make sure the mouse will be able to get the food, whether it's a big mouse or a little mouse? Why?"

Find Out condition: "Now [the next day] they decide to find out whether it's a big mouse or a little mouse. [When the Find Out condition came first, the two boxes were introduced at this point in the same way as in the Feed condition, and the Control questions were asked. When the Find Out condition came second, the experimenter explicitly contrasted the following task with the preceding one by saying 'Now they don't care anymore about feeding the mouse. What they want to know now is if it's a big mouse or a small mouse.'] Remember, if they put food into a box and in the morning the food is no longer there, then they know that the mouse came inside the box during the night and ate it."

Find Out question: "Which box should they put the food into if they want to find out if the mouse is a big mouse or a baby mouse? Why?"

The child's answer was followed by the Inconclusive Test question: "If they put the food into the box with the big opening and in the morning the food is gone, will they know if the mouse is big or small? Why? Why not?"

This was followed by the Conclusive Test question: "If they put the food into the box with the small opening and in the morning the food is gone, will they know if the mouse is big or small? What would it tell them? And if the food is still there the next morning, will they know then if the mouse is big or small? What would it tell them?"

In cases where the Find Out question was answered incorrectly but both the Inconclusive and Conclusive Test questions were answered correctly, the subject was then asked the Find Out question again.

Results and Discussion

All children answered the Control questions correctly, indicating their understanding that both a big and a small mouse could fit into the house with the large opening, whereas only a small mouse could fit into the house with the small opening.

Test Questions

Since our two studies were presented in counterbalanced order, half the subjects had just participated in Study 2. This had no effect on performance in Study 1.

Children's answers on the test questions could be classified into five response patterns (see Table 1):

Pattern 1: Correct differentiation.—For both age groups, the most frequent pattern was the correct pattern (1), that is, a clear differentiation between producing a positive result and testing a hypothesis. Children's answers fell into this pattern if they correctly answered and justified both the Feed and the Find Out questions; all children who correctly answered these two questions also correctly answered the Inconclusive and Conclusive Test questions. Answers to the Feed question were scored as correct if children justified their choice of the house with the big opening by saying that both a big and a small mouse would fit through this opening. Answers to the Find Out question were scored as correct if children justified their choice of the house with the small opening by laying out the logic of this test in one of the following ways:

Full logic (eight first graders and nine second graders): "They should take the house with the small opening, and if the food is gone, this tells them that it is a small mouse, and if it's still there it is a big mouse."

Partial logic (one first grader and one second grader): "If the food is gone, they know it is a small mouse."

Implicit logic (two second graders): "The big house would not tell whether the mouse was big or small because both could fit in."

Eleven (of 20) first graders and 12 (of 14) second graders were classified as showing the differentiated pattern (1). Nine of these 11 first graders and all 12 second graders chose the house with the small opening in response to the Find Out question and justified their choice in one of the ways specified above. One first grader did not justify his choice of the house with the small opening in response to the Find Out question but correctly answered both the Inconclusive and the Conclusive Test questions, thereby indicating that he understood the logic of testing the hypothesis. Another first grader suggested using both houses in response to the Find Out question, arguing that "if the food in the little one is still there but the food in the box with the big hole is gone, it must be a big mouse." Since he distinguished testing a hypothesis from producing a positive result, he was scored as showing the differentiated pattern. Although he did not choose the most efficient test probed for, he clearly suggested an empirical procedure for testing the hypotheses.

Incorrect patterns (2–5).—Children in these patterns answered the first Find Out question incorrectly, thus initially failing to differentiate between producing a positive effect and testing a hypothesis. Children in patterns 2–4 (35% of the first graders and 14% of the second graders) showed a tendency to produce a positive result; that is, not only did they choose the box with the big opening in the Feed condition, they also chose it in the Find Out condition, and most gave nearly identical justifications ("both could fit in here"). Three of these children, all first graders, indicated by their wrong answers to the subsequent Inconclusive Test

TABLE 1

FREQUENCIES OF RESPONSE PATTERNS IN STUDY 1

Response Pattern	F	FO	IT	CT	RFO	Grade 1 (n = 20)	Grade 2 (n = 14)
1	+	+	+	+	0	11 (55)	12 (86)
2	+	−	+	+	+	3 (15)	2 (14)
3	+	−	+	+	−	1 (5)	0
4	+	−	−	+	0	3 (15)	0
5	−	−	−	+/−	0	2 (10)	0

NOTE.—Percentages are in parentheses. F = Feed question, FO = Find Out question, IT = Inconclusive Test question, CT = Conclusive Test question, RFO = Repeat Find Out question, + = correct answer and justification, − = incorrect answer, 0 = not asked. Pattern 1: correct differentiation between producing an effect and testing a hypothesis; Pattern 2: initial nondifferentiation, recognition of inconclusiveness, correct choice of conclusive test on repeated probing; Pattern 3: persistent nondifferentiation despite recognition of inconclusiveness; Pattern 4: nonrecognition of inconclusiveness; Pattern 5: Pervasive nonunderstanding.

question that they did not understand the ambiguity of the outcome of the inconclusive test (Pattern 4). Six children, four first graders and two second graders, did correctly answer the Inconclusive Test question; of these, all but one (a first grader) correctly answered the Repeat Find Out question as well, choosing the box with the small opening and fully or partially elaborating the logic of this test (Pattern 2). One first grader answered the Inconclusive Test question correctly but answered "don't know" to the Repeat Find Out question (Pattern 3).

Two first graders' responses did not conform to any of the above patterns. They answered all (or all but one) questions incorrectly (Pattern 5); they chose the house with the small opening in response to the Feed question, giving irrelevant justifications; in response to the Find Out question, one of them chose the house with the big opening and one chose both houses.

Pattern 1 (differentiation) was not significantly more frequent in second than in first graders, Fisher's exact probability test, $p = .063$. First graders showed the differentiated response pattern more often when the Find Out condition preceded than when it followed the Feed condition; the order effect, however, just failed to reach significance, Fisher's exact probability test, $p = .056$. Note that six of the seven first graders who incorrectly gave a "positive effect" response to the Find Out question received the Feed condition first, and five of them merely repeated their previous responses to the Feed question. First graders, then, may be more likely than second graders to misinterpret the task of testing a hypothesis as one of producing a desirable effect when a preceding task biases them toward such an interpretation.

Seventy percent of the first graders and all second graders eventually (i.e., upon the initial [Pattern 1] or the Repeat Find Out question [Pattern 2]) chose a conclusive test, second graders being more successful than first graders, Fisher's exact probability test, $p = .03$. Second graders were not significantly more successful than first graders on the Inconclusive Test question, Fisher's exact test, $p = .055$. If only those children were taken into account who correctly answered the Inconclusive Test question (15 first graders [75%] and 14 second graders [100%]), then there was no significant difference between first and second graders' performance on either the initial or the Repeat Find Out questions (Fisher's exact test; $p = .22$, first question, $p = .51$, second question). Thus, first graders who understood that an inconclusive test would not allow one to decide between two conflicting hypotheses did not seem to have an additional problem in answering the Find Out question.

First- and second-grade children showed considerable competence in choosing a conclusive empirical test to decide between two simple alternative hypotheses. More than half of the first graders and a large majority of the second graders were correct on the initial Find Out question. This performance is very impressive, considering that this question required children to anticipate what follows from a conclusive and an inconclusive test for the protagonists' knowledge of the relevant fact and to choose a test on the basis of this understanding. Thus, young elementary school children seem not only to be able to distinguish between warranted and unwarranted inference but also to understand the implications of this distinction for the logic of hypothesis testing. Not surprisingly, 25% of the first graders did not understand the inconclusiveness of one of the proposed tests. This is consistent with the performance of 6–7-year-old children on similar tasks requiring the recognition of indeterminacy (e.g., Somerville et al., 1979). However, virtually all of the children who did recognize indeterminacy were eventually able to choose an adequate test for a hypothesis. This indicates that they did not have a fundamental problem in understanding the notion of "testing a belief" by means of gathering indirect evidence.

Study 2

If young children understand what it means to test ideas, they should be able to spontaneously generate a method for putting an idea to an empirical test. Previous research (e.g., Dunbar & Klahr, 1989) indicates that elementary school children do not design well-formed experiments. Nevertheless, even if they fail to generate perfectly conclusive tests, they may succeed in setting up a condition that provides relevant information; that is, they may have a basic understanding of experimentation. Children's spontaneous attempts to generate tests of simple hypotheses in domains they understand should provide evidence of this basic understanding. In Study 2, therefore, prior to presenting a forced choice between a conclusive and an inconclusive test we asked children how they would find out which of two conflicting hypotheses is right. Study 1 assessed children's ability to choose between alternative hypotheses in an artificially contrived world (only one mouse, no possibilities to catch it and look at it, etc.). This situation is very unlike scientific hypothesis testing. In Study 2 we posed children a genuine scientific problem, namely, how to find out whether an unfamiliar animal has a good or a poor sense of smell. Here an indirect experimental test is the only way to decide between alternative hypotheses since the animal's sense of smell is not observable.

Method

Subjects

As mentioned above, to provide information about within-child consistency, we tested the same subjects who participated in Study 1. The children received the two tasks in counterbalanced order.

Procedure

Each child was presented with the following story, accompanied by illustrations. "This is a story about Tom and Mike who got a new pet animal. It is a very rare animal called an aardvark. They want to know a lot about their pet. One thing they would really like to know is how well this animal can smell. They know that some animals have very sensitive noses, much better than people, so that they can smell things that we can't smell. And other animals have poor noses just like people so that they can only smell things that have a really strong smell. Tom says an aardvark is much like a pig and pigs have very good noses, so he thinks that an aardvark also has a very sensitive nose. Mike says an aardvark is also very similar to a sheep and sheep have poor noses, so he thinks that the aardvark does not smell very well."

Spontaneous Test question: "They want to find out who is right. What could they do to find this out?"

After the child had given an answer, the experimenter proceeded: "Tom and Mike have an idea. They want to put a piece of food in the animal's box and cover it up with sand so that the aardvark can't see it and then they want to see whether the aardvark finds it or not. They have two different kinds of food. One has a very strong smell and the other one has very little smell. The piece with the strong smell smells so strongly that even people can smell it, so an animal with a weak nose can smell it, and an animal with a good nose can, of course, smell it too. The smell of the piece with the weak smell is so weak that we can't smell it, so an animal with a bad nose won't be able to smell it, but an animal with a good nose will be able to smell it."

Children were then asked what outcome the two story figures would predict for each of the pieces of food (Prediction questions):

"Remember, Tom thinks that the aardvark has a very good nose. Will he think that it will find the strong smelling food? Will he think that it will find the weak smelling food?"

"Remember, Mike thinks that the aardvark has a very poor nose. Will he think that it will find the strong smelling food? Will he think that it will find the weak smelling food?"

If a child made an incorrect prediction, the experimenter pointed out the relevant facts again (e.g., the strong smelling food has such a strong smell that even people who have poor noses can smell it) and asked the question again.

When the child had answered all Prediction questions, the experimenter proceeded as follows: "Now they want to do an experiment. They want to find out whether the aardvark has a good nose or a bad nose by putting a piece of food out in the sand and seeing if the aardvark finds it."

Find Out question: "To find out whether Tom is right, who thinks that the aardvark has a very good nose, or whether Mike is right, who thinks that the aardvark has a very poor nose, what should they do? Which piece of food should they hide to find out whether the animal has a good nose or a poor nose? Why?"

Inconclusive Test question: "Suppose they put the piece of food with the strong smell in the box and the animal found it. Would they then know who is right or wouldn't they know? Why?"

Conclusive Test question: "Suppose they put the piece with very little smell in the box and the animal found it. Would they then know who is right or wouldn't they know? What would it tell them? And if the animal did not find it, would they then know who is right or wouldn't they know? What would it tell them?"

If the Find Out question was answered

incorrectly but both the Inconclusive and the Conclusive Test questions were answered correctly, the subject was then asked the Find Out question again.

Results and Discussion

Spontaneous Test Question

In response to the Spontaneous Test question, three different types of answers were given: a minority of both first and second graders (25% and 21%, respectively, see Table 2) spontaneously generated a conclusive test; that is, they explicitly specified a condition that would allow only an animal with a good sense of smell to find the food (e.g., "put a piece of food very far away; if it can smell it it has a good nose, if it can't it has a bad nose." "Put something out that is very hard to smell for us; if he can find that, he has a good sense of smell"). Thirty-five percent of the first graders and 57% of the second graders gave answers that were classified as "inconclusive tests." These tests involved measuring the animal's sense of smell by determining whether it could find a particular kind of food. They were classified as inconclusive because they either did not specify a relevant property of food, for instance, the strength of its smell or its distance from the animal (e.g., "put out something with a good smell like chocolate and see if it can find it") or they did not elaborate the logic of a relevant test (e.g., "put out pieces of food with a strong smell and pieces with a weak smell"). A third type of answer did not involve an empirical test at all but rather suggestions to ask experts or to make observations (e.g., "ask at a pet store," "look it up in an encyclopedia," "watch the animal"). Fifteen percent of the first graders and 21% of the second graders gave this type of answer. Five first graders (25%) gave no answer or an irrelevant answer.

Notice that our Spontaneous Test question did not explicitly request an empirical test; it merely asked subjects what the protagonists could do to find out who was right. An appeal to experts is a perfectly good strategy for deciding between the alternative hypotheses. Nevertheless, since such a strategy does not involve empirical tests of hypotheses, it cannot count as evidence of children's ability to spontaneously generate such tests. It is perfectly plausible that some of these children, too, could have devised an empirical test if it had been required by the task or if an appeal to experts were not possible. Indeed, subjects who had just participated in Study 1, and therefore had just decided between two empirical tests, were slightly less likely to appeal to experts and more likely to actually generate their own empirical test, Fisher's exact test, $p = .043$, one-tailed. However, even among those children who received Study 2 first, 50% spontaneously suggested an empirical test. In view of this, children's responses to this question must be taken as a conservative measure of their ability to generate empirical tests.

There were no significant differences between first and second graders on the Spontaneous Test question. The majority of the children spontaneously devised an empirical test to decide between the competing hypotheses, although only a minority explicitly specified a conclusive test.

Prediction Questions

All children except for two first graders were able to derive correct predictions from information about the story characters' hypotheses and about the properties of the food. Seven first graders and two second graders needed an additional explanation on one of the four Prediction questions (usually the experimenter had to clarify that the

TABLE 2

Number of Children Who Spontaneously Suggested Conclusive and Inconclusive Empirical Tests in Study 2

	Grade	
Response Category	1 (n = 20)	2 (n = 14)
Empirical test: Conclusive	5 (25)	3 (21)
Empirical test: Inconclusive	7 (35)	8 (57)
Ask expert, watch	3 (15)	3 (21)
No response, irrelevant	5 (25)	0

Note.—Percentages in parentheses.

strong smelling piece of food was so strong that one could smell it even if one had a poor sense of smell). The two first graders who were unable to derive correct predictions also failed to answer the subsequent Find Out and Inconclusive Test questions correctly.

Test Questions

There was no effect of order of presentation of the two studies on responses to these questions.

Pattern 1: Correct.—As in Study 1, the majority of children of both age groups gave correct answers to the Find Out and the Conclusive and Inconclusive Test questions (60% of first graders and 93% of second graders; see Table 3).

Of these children, all the second graders and two-thirds of the first graders also correctly justified their choice of the weak smelling piece of food: six first graders and nine second graders justified their answer by fully elaborating the logic of the test (e.g., "They should take a weak smelling piece because if he has a poor nose he won't find it and if he has a good nose he will find it"); two first graders and two second graders elaborated this logic partially (e.g., "If he has a poor nose he won't find it"); and two second graders gave an implicit justification ("They should take the weak one, because both an animal with a good nose and an animal with a bad nose would find the strong smelling piece"). The other four first graders who gave the correct response pattern did not justify their correct answer to the Find Out question; they did, however, correctly answer and justify the subsequent Inconclusive Test question and correctly specify the logic of a conclusive test in response to the Conclusive Test question.

Incorrect patterns.—Four first graders (20%) and one second grader (7%) chose the strong smelling piece of food in response to the Find Out question (Patterns 2 and 3); two of these children (one first grader and one second grader) gave "positive result" justifications ("so that the animal can find it"). The other three children's justifications were not readily interpretable ("take the strong piece because it smells so strong"). Only one of the children who chose the strong smelling piece answered the subsequent Inconclusive Test question correctly; this child then also chose the correct test on the Repeat Find Out question (Pattern 2). The other four children did not recognize the inconclusiveness of testing the strong smelling piece of food; that is, they said that the story figures would know that the animal had a good nose if it found the strong smelling piece (Pattern 3).

The "other" pattern included four first graders (20%) who either gave "don't know" responses to the Find Out question or who chose the weak smelling piece but gave irrelevant justifications and then answered the Inconclusive Test question incorrectly.

The correct response pattern (Pattern 1) was significantly more frequent in second graders than in first graders, Fisher's exact probability test, $p = .036$, with second graders performing almost perfectly. However, even among first graders, 60% were correct on the first Find Out question and 65% were correct eventually (Repeat Find Out question).

Seven of the eight children who had spontaneously designed a conclusive test also correctly answered the forced-choice Find Out question. The one child who answered incorrectly (a first grader) had spontaneously suggested manipulating the dis-

TABLE 3

FREQUENCIES OF RESPONSE PATTERNS IN STUDY 2

Response Pattern	FO	IT	CT	RFO	Grade 1 (n = 20)	Grade 2 (n = 14)
1	+	+	+	0	12 (60)	13 (93)
2	−	+	+	+	1 (5)	0
3	−	−	+	0	3 (15)	1 (7)
Other					4 (20)	0

NOTE.—Percentages are in parentheses. FO = Find Out question, IT = Inconclusive Test question, CT = Conclusive Test question, RFO = Repeat Find Out question; correct pattern = Pattern 1.

tance of the food from the animal ("put food far away") and did not respond to the experimenter's suggestion of another method of testing the hypothesis. The majority of the children who spontaneously suggested an inconclusive test (12 of 15) correctly chose the conclusive test in response to the Find Out question, whereas only six of 11 children in the "ask expert/watch" or "no response" category did so. The difference in performance on the forced-choice task between those children who had suggested an empirical test spontaneously and those who did not approached significance, Fisher's exact test, $p = .09$.

Within-child consistency between Study 1 and Study 2 was very high: 28 of the 34 children were either correct on both tasks (21) or incorrect on both (7), phi coefficient $= .51$, $p < .01$. Of the six children (three in each age group) who responded correctly on only one task, four succeeded in Study 2 and two in Study 1.

Study 2 replicated the findings of Study 1, indicating that young elementary school children can distinguish between a conclusive and an inconclusive test of a simple hypothesis. Thirty-five percent of the first graders failed to understand the inconclusiveness of one of the proposed tests. Not surprisingly, these children failed the task of choosing a conclusive test. However, in both first and second graders, all the children who did understand the inconclusiveness of testing the animal's sense of smell with a strong smelling piece of food chose the right experiment to decide between the two competing hypotheses. Furthermore, all but one of these children did so on the initial Find Out question, before their understanding of the logic of the test was probed.

In Study 2, we did not explicitly contrast the task of hypothesis testing with a task of producing a desirable effect. Under these conditions, we found the latter tendency in only two (of 34) children, even though feeding a pet is intrinsically desirable. This suggests that children do not spontaneously interpret the task of hypothesis testing as one of generating desirable effects. Rather, our results suggest that most of the children who chose the strong smelling piece of food to test the animal's sense of smell incorrectly believed that a positive outcome of this test would allow them to conclude that the animal had a strong sense of smell. This shows that they did not understand the inconclusiveness of this test, although they may very well have understood the idea of "testing" itself.

Moreover, Study 2 showed that even first-grade children have spontaneous ideas about testing hypotheses. The majority of the children suggested an empirical test that was clearly relevant to the present problem, although only a minority spontaneously elaborated a fully conclusive test. It should be noted, however, that our scoring system was conservative, so that children who suggested manipulating a relevant variable (strength of smell, distance) but did not elaborate the logic of this manipulation were scored as suggesting an inconclusive test. We conclude that young elementary school children have the notion of putting ideas to an empirical test and do not confound testing ideas with generating positive effects.

General Discussion

Our studies indicate that young elementary school children differentiate hypothetical beliefs from evidence: The majority of first graders and almost all second graders demonstrated that they understood the goal of testing a hypothesis as opposed to the goal of producing a positive effect. Furthermore, they were able to distinguish between a conclusive and an inconclusive test for simple hypotheses about determinate states of affairs and understood the inferences that could be made from the outcome of a conclusive test. Many children were even able to spontaneously generate a procedure for gathering indirect evidence relevant to a belief. However, it must be noted that few children of these ages spontaneously devised a conclusive experiment, even though they recognized one when we laid it out for them. It is an open question at what age children would solve this problem spontaneously.

As we argued in the introduction, a demonstration of children's ability to explicitly lay out the inferential relation between hypotheses and evidence provides clear support for the position that they do differentiate the notions of "hypothesis" and "evidence." Our studies demonstrate this ability in young elementary school children. These findings are consistent with studies showing that children of this age understand that beliefs can be acquired through inference (Sodian & Wimmer, 1987), and that they can distinguish between warranted and unwarranted inference (Somerville et al., 1979). Although some first graders had difficulty understanding that the outcome of an inconclusive test would not allow one to decide between two hypotheses, the majority distinguished between the epistemic consequences of conclusive and inconclusive tests. In addition to what was known from previous research, our studies show that most of the first graders and the large majority of the second graders were not only able to elaborate the epistemic consequences upon request but were able to anticipate them in their initial choice of a conclusive test. Thus, young children are not only able to recognize indeterminacy, but they are aware of its consequences for the testing of beliefs. Furthermore, virtually all children of both age groups who understood the indeterminacy of one of the proposed tests recognized the consequences of this indeterminacy for the choice of an adequate empirical test upon repeated probing. Roughly 70% of the first graders and all second graders eventually chose a conclusive test.

The aim of our studies was to test whether elementary school children have an undifferentiated concept of "hypothesis" and "evidence," as has been proposed in the scientific thinking literature. Our findings indicate that this is not the case. However, our results do not imply that children of this age understand all relevant aspects of the logic of hypothesis testing, especially of testing hypotheses about causal relations between variables, as have been typically studied in tasks that diagnose scientific thinking. These tasks differ from ours in a number of important ways.

First, an important part of scientific activity that was probed for in the literature on scientific thinking is *generating* hypotheses about phenomena. Children were usually first shown an effect that was salient and engaging to produce (e.g., making a robot work [Dunbar & Klahr, 1989], making chemicals change color [Kuhn & Phelps, 1982], making a yeast reaction occur [Carey, Evans, Honda, Jay, & Unger, 1989]) and were asked to explore what causes this effect. In contrast, we did not present children with an effect that could have been caused by a number of factors. Rather, we presented them with two different beliefs about a state of affairs and required them to reason about evidence that would allow a decision between these beliefs. Thus, we separated the task of generating hypotheses from that of testing known hypotheses. Under these conditions, the majority of the children were clearly able to distinguish between the task of testing a hypothesis and that of producing an effect. The minority of the children who showed a "positive effect orientation," that is, who in Study 1 suggested choosing the house with the big opening to determine the size of the mouse "because then both a big and a small mouse would fit in (and get the food)," had (with one exception) received the "feed" task (in which they were explicitly asked to produce a positive result) before the hypothesis-testing task. Thus, they seemed to misinterpret hypothesis testing as producing an effect only when this interpretation was previously suggested to them. The finding by Kuhn and Phelps (1982) and Dunbar and Klahr (1989) that children were trying to produce effects rather than to understand phenomena may reflect their difficulties in generating hypotheses in a new domain. In generating hypotheses, it is of course not irrational to begin an exploration of a phenomenon by replicating a salient effect. Note that Karmiloff-Smith and Inhelder (1974/75) observed that most of their subjects began with a "positive result orientation," but only the youngest children (preschoolers) maintained that orientation throughout the whole session; the older children (elementary school age), on the other hand, quickly moved on to construct and test theories.

Second, an important requirement in testing scientific hypotheses is to actively seek disconfirmation. It is well known that even adults often do not seek disconfirming evidence when testing hypotheses (e.g., Wason, 1960). Karmiloff-Smith and Inhelder (1974/75) found that their subjects did not actively seek counterevidence to their theories about block balancing. However, the elementary school children were able to recognize counterevidence when it was presented to them. In our studies, children were asked to choose a test to decide between two alternative hypotheses. Thus, they did not have to generate an alternative to an existing hypothesis and seek evidence that would support this alternative hypothesis.

Third, the relation between hypotheses and data in typical scientific thinking tasks is less obvious than in our studies. In our studies, the outcome of a single test (e.g., "cheese gone from the house with the small opening") allowed a definite conclusion about the true state of affairs (cause of the disappearance of the cheese was the presence of a mouse who could fit through the small opening). Such conclusions can never be drawn from covariation data obtained in testing hypotheses about causal relations between variables. Patterns of covariation do not allow one to infer causation; they just make one or another causal hypothesis more plausible. Young children may understand how to devise a conclusive test of a hypothesis before they begin to understand which patterns of data support a hypothesis. In the simplest case, a hypothesis about whether or not a variable affects another requires at least two tests: To test, for instance, whether it is plausible to assume that large tennis balls are better for a game than small ones, one has to test both large and small balls, while keeping all other factors constant. Thus, an understanding of hypothesis testing as investigated in Kuhn et al.'s tasks requires children to understand that a single test is not sufficient to draw a conclusion. In contrast, in our tasks, there was a single conclusive test that had to be distinguished from an inconclusive one.

Fourth, children (like scientists) may have difficulty revising beliefs that are held with some conviction. Our studies show that young children understand in principle how evidence bears on the truth or falsity of hypotheses. The nature of our tasks made it implausible that they should have any previous biases about the truth of either of the two conflicting hypotheses. In contrast, in the scientific thinking literature children's understanding of experimentation was often tested in contexts where they had strong favored hypotheses of their own. Such hypotheses may have been grounded in evidence from previous experience, and children may have had difficulty understanding that such beliefs are open to disconfirmation. In fact, many children seemed not even to realize that there could be alternative hypotheses to their own.

In sum, there are many dimensions along which children's understanding of scientific reasoning may develop. These include an understanding of the necessity of seeking disconfirming evidence, and a full appreciation of the logic of testing hypotheses about causal relations between variables. Our studies show that young elementary school children possess the metaconceptual basis for acquiring this knowledge: They distinguish belief and evidence, and they know how to put hypotheses to the test.

References

Braine, M. D. S., & Rumain, B. (1983). Logical reasoning. In J. H. Flavell & E. M. Markman (Eds.), P. H. Mussen (Series Ed.), *Handbook of child psychology: Vol. 3. Cognitive development* (pp. 266–340). New York: Wiley.

Bullock, M. (1989). *Hypothesis testing in elementary school children.* Paper presented at the biennial meeting on Developmental Psychology in Germany, Munich.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). "An experiment is when you try and see if it works": A study of seventh grade students' understanding of the construction of scientific knowledge. *International Journal of Science Education,* **11,** 514–529.

Case, R. (1974). Structures and strictures: Some functional limitations on the course of cognitive growth. *Cognitive Psychology,* **6,** 544–573.

Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery processes. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 109–144). Hillsdale, NJ: Erlbaum.

Dunbar, K., Klahr, D., & Fay, A. (1989). *Developmental differences in scientific reasoning processes.* Paper presented at the biennial meeting of the Society for Research in Child Development, Kansas City, Mo.

Fabricius, W. V., Sophian, C., & Wellman, H. M. (1987). Young children's sensitivity to logical necessity in their inferential search behavior. *Child Development,* **58,** 409–423.

Haake, R. J., & Somerville, S. C. (1985). Development of logical search skills in infancy. *Developmental Psychology,* **21,** 176–186.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic.

Karmiloff-Smith, A., & Inhelder, B. (1974/75). If you want to get ahead, get a theory. *Cognition,* **3,** 195–212.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review,* **96,** 674–689.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills.* New York: Academic Press.

Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. In H. Reese (Ed.), *Advances in child development and behavior* (Vol. **17,** pp. 1–44). New York: Academic Press.

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

Pieraut-LeBonniec, G. (1980). *The development of modal reasoning: Genesis of necessity and possibility notions.* New York: Academic Press.

Robinson, E. J., & Whittaker, S. J. (1985). Children's responses to ambiguous messages and their understanding of ambiguity. *Developmental Psychology,* **21,** 446–454.

Siegler, R. S., & Liebert, R. M. (1975). Acquisition of formal scientific reasoning by 10- and 13-year-olds: Designing a factorial experiment. *Developmental Psychology,* **11,** 401–402.

Siegler, R. S., Liebert, D. E., & Liebert, R. M. (1973). Inhelder and Piaget's pendulum problem: Teaching adolescents to act as scientists. *Developmental Psychology,* **9,** 97–101.

Sodian, B. (1988). Children's attributions of knowledge to the listener in a referential communication task. *Child Development,* **59,** 378–385.

Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development,* **58,** 424–433.

Somerville, S. C., Hadkinson, B. A., & Greenberg, C. (1979). Two levels of inferential behavior in young children. *Child Development,* **50,** 119–131.

Sophian, C., & Somerville, S. C. (1988). Early developments in logical reasoning: Considering alternative possibilities. *Cognitive Development,* **3,** 183–222.

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development,* **51,** 1–10.

Wason, P. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology,* **12,** 129–140.

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