The Development of Reasoning about Causal and Noncausal Influences on Levers

Eric Amsel
Vassar College

Geoff Goodman
Teachers College, Columbia University

Dallas Savoie and Megan Clark
University of Saskatchewan

AMSEL, ERIC, GOODMAN, GEOFF, SAVOIE, DALLAS; and CLARK, MEGAN *The Development of Reasoning about Causal and Noncausal Influences on Levers. Child Development*, 1996, **67**, 1624–1646. 2 studies examined 5–12-year-olds' judgments regarding the behavior of balance scales and other levers whose arms varied in a causal (the number of equally weighted objects or their distances from the fulcrum) or a noncausal (the color, position, or orientation of objects) variable. There were age-related increases in correct judgments for each causal and noncausal variable, with children tending to make correct judgments about the influence of physical features of objects (their number and color) at an earlier age than they did about spatial relations between objects (their distance, orientation, and position). Children's patterns of errors judging the influence of causal (particularly distance) and noncausal (particularly position and orientation) variables were different, and there was no relation between children's correct judgments regarding causal and noncausal variables The results suggest that there are separate processes underlying children's ability to identify causal and dismiss noncausal influences on levers which are dependent on the kinds of features (physical or both physical and spatial) which children conceive of as potentially influencing the behavior of levers.

Children have been credited with the ability to form and revise intuitive theories of the world (Wellman & Gelman, 1992). Intuitive theories are defined as integrated networks of explanatory concepts and causal beliefs regarding phenomena in domains such as psychology (Gopnik & Astington, 1988; Gopnik & Wellman, 1992; Perner, 1991), biology (Carey, 1985a; Gelman & Markman, 1986; Keil, 1992), and physics (Carey, 1991; Smith, Carey, & Wiser, 1985; Vosniadou & Brewer, 1992). It has been claimed that intuitive theories have the same predictive and explanatory functions for children that formal theories have for scientists (Brewer & Samarapungavan, 1991; Gopnik & Wellman, 1992; Wellman & Gelman, 1992). Moreover, in response to limitations, gaps, or conflicts in their predictions and/or explanations of phenomena, children and scientists alike are thought to revise their theories (Brewer & Samarapungavan, 1991; Carey, 1985a, 1985b, 1991; Gopnik & Wellman, 1992). The result is a revised theory of greater predictive and explanatory power and adequacy than the initial one.

The data for Study 1 were collected while the first author was a Social Sciences and Humanities Research Council of Canada Postdoctoral Fellow at Yale University and the collection of data for Study 2 was supported by a University of Saskatchewan President's grant and Natural Sciences and Engineering Research Council of Canada (NSERC) operating grant to the first author, and by an NSERC summer research grant to the third author Portions of the results were presented at the meeting of the Society for Research in Child Development, 1987, Baltimore, and at the Conference of the Cognitive Science Society, 1991, Chicago. The authors would like to thank the faculty, staff, and students of St. Francis Xavier School in New York City, where Experiment 1 was run, and the Protestant School Board of Saskatoon, Saskatchewan, in whose schools Study 2 was run Colleagues and students at Yale University and the University of Saskatchewan have been very helpful in commenting on and critiquing the research. Helpful comments on earlier versions of this manuscript were made by Judi Amsel, Merry Bullock, David Moshman, and three anonymous reviewers. Geoff Goodman is now at Cornell Medical Center. Megan Clark is now at the University of British Columbia. Reprint requests should be sent to the first author at Department of Psychology, Weber State University, 1202 University Circle, Ogden, UT 84408-1202 (Eramsel@cc.weber edu).

[*Child Development*, 1996, **67**, 1624–1646. 0009-3920/96/6704-0013$01 00]

As evidence that children revise a theory, researchers cite the breadth, synchrony, and coherence of changes in their reasoning about related phenomena. *Breadth* refers to a range of changes in children's reasoning about related phenomena, reflecting a restructuring of an entire network of beliefs and concepts, rather than to a simple change of a particular belief or concept about a specific phenomenon. *Synchrony* refers to the fact that the changes in reasoning occur at roughly the same times, reflecting a restructuring of a network of beliefs and concepts in response to a perceived weaknesses in the initial theory.[1] Finally, *coherence* refers to the fact that the changes in reasoning are directed toward increasing the explanatory and predictive power and adequacy of the revised theory over the initial one. For example, as evidence that young children's theory of astronomy undergoes revision, Vosniadou and Brewer (1992) demonstrated that parallel age-related increases (synchrony) in 6–11-year-olds' correct responses to questions about a range of astronomical phenomena (breadth) were based on progressively more adequate models of the earth (coherence). Similarly, as evidence that young children revise a theory of mind, Gopnik and Astington (1988) cite that parallel age-related increases (synchrony) in 3–5-year-olds' correct performance on a range of tasks assessing their understanding of mental representations (breadth) were correlated with each other, reflecting their acquisition of a representational theory of mind (coherence).

While there may be broad, synchronous, and coherent changes in children's reasoning, reflecting the revision of a theory, very little is known about the processes underlying theory revision itself. The present research addresses the question of the processes of theory revision in children and in particular how age-related differences arise in causal and noncausal judgments about theory-related phenomena. There is wide agreement that theory revision involves age-related differences in children's judgments about causal and noncausal influences on theory-related phenomena (Carey, 1985a, 1985b; Gopnik & Wellman, 1992; Wellman & Gelman, 1992). For example, in acquiring a representational theory of mind, children come to identify mental (e.g., mistaken beliefs about true states of the world) influences on an agent's behavior and dismiss nonmental (e.g., true states of the world) influences (Perner, 1991). However, there has been only limited discussion about how age-related differences in children's causal and noncausal judgments arise in the context of theory revision.

Two accounts of the process(es) underlying such age-related differences in children's judgments seem plausible. On the one hand, theory revision may involve changes en masse in children's judgments about causal and noncausal influences on theory-related phenomena. The changes in causal and noncausal judgments may occur together and in a related fashion as children acquire new causal knowledge of a domain. Each time a new causal understanding of a domain is acquired, there is a corresponding change in children's judgments of causal and noncausal influences on theory-related phenomena. Such a view of theory revision is supported by research on the development of causal reasoning which suggests that as children acquire knowledge about the mechanism by which particular phenomena are produced, they are better able to both correctly identify causal and dismiss noncausal influences on those phenomena (Bullock, 1985; Bullock, Gelman, & Baillargeon, 1982; Shultz, 1982).

On the other hand, the revision of a theory may involve *piecemeal* changes in children's causal and noncausal judgments regarding theory-related phenomena. In this view, children revise a causal understanding of a domain as a result of independent changes in their causal and noncausal judgments of the domain. Independent changes in children's judgments may occur as separate processes for identifying causal and dismissing noncausal influences of phenomena in a domain are initiated in response to perceived limits or gaps in children's previous causal understanding of the domain. Support for this account of theory revision comes from research on the development of scientific reasoning which suggests that judging variables as causal (inclusion) or noncausal (exclusion) involves independent inferential skills, with young children perhaps more able to do the former than the latter (Inhelder & Piaget, 1958; Kuhn, Amsel, & O'Loughlin, 1988; Kuhn, Schauble, & Mila-Garcia, 1992; Schauble, 1990).

[1] There is no requirement that the changes in reasoning occur suddenly. Indeed, as some theorists have noted, children's theory revision is slow and gradual, occurring over a period of years (Carey, 1985b, 1991; Smith et al., 1985; Vosniadou & Brewer, 1992).

In the present research, we examined if there are age-related differences in children's judgments about causal and noncausal influences on phenomena reflecting the revision of a theory and whether or not such differences occur en masse, in an interrelated manner, or in a more piecemeal and unrelated fashion. In particular, we examined whether there are not only broad, synchronous, and coherent but also related (or unrelated) differences with age in children's correct judgments about causal and noncausal influences on the behavior of levers. Levers are simple machines that are composed of a rigid bar that moves about a support or fulcrum. Levers behave by an "effort" (e.g., a push, pull, or mass), applied to a point on the lever, balancing, raising, or overcoming a "load" (e.g., a mass or other type of resistance) at another point. For example, balance scales (e.g., pan balances or balance beams) are a class of levers with the fulcrum placed between two masses serving as (arbitrarily) the load and the effort, such that the two forces are in equilibrium. A noncausal variable is a feature which varies between the load and effort but has no influence on the behavior of levers, whereas a causal variable is a feature varying between the effort and load that does influence the behavior of levers. Two causal variables are the *magnitude* of the effort and load and their *distance* from the fulcrum. For example, all other things being equal, a balance scale falls to the side of the arm with more equally weighted objects (i.e., the magnitude of the force) or with objects that are further away from the fulcrum (i.e., the distance from the fulcrum). The two causal variables can be coordinated in order to gain a mechanical advantage such that an effort of a lesser magnitude than the load but applied farther away from the fulcrum than the load can nonetheless balance, raise, or overcome the load.

Previous research has shown that there are developmental differences in children's identification of the influence of causal variables on the behavior of balance scales. Five- and 6-year-olds tend to identify the magnitude of forces (as measured by the *number* of equally weighted objects) as a causal variable influencing the behavior of balance scales, but not the *distances* of these objects from the fulcrum (Case, 1985; Ferretti & Butterfield, 1986; Ferretti, Butterfield, Cahn, & Kerkeman, 1985; Inhelder & Piaget, 1958; Klahr & Siegler, 1978; Siegler, 1976, 1978, 1981, 1985; Siegler & Klahr, 1982; Siegler & Taraban, 1986; Surber & Gzesh, 1984; Wilkening & Anderson, 1982, 1991). For example, these children judge that the arms of balance scales fall to the side of the arm with a greater number of equally weighted objects (blocks, weights, etc.) and balance if the number of objects on both arms is the same, irrespective of the distance of objects from the fulcrum. Indeed, the children often fail to note differences in the distances of objects on the arms of balance scales when asked to simply copy the arrangement of objects (Siegler, 1976). However, by age 10 or so, children tend to correctly encode and predict the influence of both the number of objects and their distance from the fulcrum, although whether these children also coordinate number and distance information remains controversial (Siegler, 1976; Wilkening & Anderson, 1991).

To test whether the differences in children's reasoning about balance scales involve the revision of a theory, we assessed their judgments of the influence of both causal and noncausal variables on the behavior of balance scales (Study 1) and other types of levers (Study 2). We hypothesized that there would be parallel increases (synchrony) in 5–12-year-olds' correct judgments regarding the influence of causal and noncausal variables on different types of levers (breadth), reflecting a growth in their understanding of why and how efforts balance, raise, or overcome loads (coherence). Extending previous research, we predicted that children should have the same difficulties identifying the influence of causal variables on the behavior of other levers that they have on balance scales. In addition, we predicted that children's misunderstanding of how levers work should lead them to have difficulty dismissing the influence of noncausal variables on their behavior. The latter hypothesis was based on findings that children fail to dismiss the influence of noncausal variables on other physics tasks, reflecting their causal misunderstanding of the domain of those tasks. For example, children often fail to dismiss the influence of the weight of objects on their free-fall rate (Selman, Krupa, Stone, & Jacquette, 1982), rate of descent on an inclined plane (Howe, Tolmie, & Rodgers, 1992; Inhelder & Piaget, 1958), and rate of completing a period on a pendulum (Inhelder & Piaget, 1958; Siegler, Liebert, & Liebert, 1973), suggesting that they misunderstand the force of gravity (Reiner, Chi, & Resnick, 1988). Indeed, children tend to judge that the weight of objects influences how fast they fall, roll, and swing, reflecting a false belief about the causal status of the variable: Instead of dismissing the weight of objects as a noncausal variable, they identify it as a causal one.

There is some preliminary evidence to suggest that young children may also have difficulty dismissing the influence of noncausal variables on the behavior of balance scales. Ausubel and Schiff (1954) found that younger children learned in an equal number of trials that a rigged balance scale fell to the longer side as they learned that it did to the side of red blocks. However, older children learned the influence of length faster than color, suggesting a more advanced understanding of the possible causal or noncausal status of these features. Metz (1993) found that some young children attribute a lack of equilibrium in a pan balance scale to other features of the objects in the pans (their color, size, or number) rather than to their weight.

The present study extends this previous research in order to more adequately examine developmental differences in children's judgments of the influence of noncausal variables on the behavior of levers. Children judged the influence of a range of noncausal variables, including the color of equally weighted objects placed on the arms of levers, their horizontal or vertical orientation, and their position above or below the arms.[2] Changing the orientation and position of objects on the arms of levers alters the spatial arrangement between objects and not merely a physical feature such as their color. In addition, the nature of children's incorrect judgments about the influence of noncausal variables on levers was examined. As with their judgments of the influence of weight on gravity-related tasks, young children may judge that noncausal variables influence the behavior of levers, reflecting false causal beliefs about the variables. For example, children may judge that balance scales always fall to the side of the arm with a particular level of a noncausal variable. Alternatively, children's judgments of noncausal variables may reflect neither a causal nor a noncausal influence, but confusion or uncertainty about the status of the variables. For example, children may make balance and imbalance judgments regarding balance scales reflecting no discernible belief regarding the causal or noncausal influence of the variable.

Finally, besides examining age-related differences in children's judgments about each noncausal variable, we assessed whether or not the differences in correct judgments about noncausal and causal variables were correlated. The correlation analysis was performed to in order to test theories of how age differences in causal and noncausal judgments arise. If the differences in children's correct judgments of the influence of causal and noncausal variables occur en masse, in a related manner, as children acquired more adequate causal knowledge about how levers work, their tendencies to correctly judge the influence of causal and noncausal variables should be correlated with each other. Children with inadequate causal knowledge should incorrectly judge the influence of causal and noncausal variables and those with adequate knowledge should correctly judge the influence of both variables. Alternatively, if the age differences in children's judgments of the influence of causal and noncausal variables occur in a more piecemeal fashion, as separate processes are initiated to infer the causal and noncausal status of variables, then children's tendencies to correctly judge the influence of causal and noncausal variables should be uncorrelated. Under this view, children may tend to correctly identify causal variables independently of and perhaps prior to dismissing noncausal ones since inclusion is an inference skill that is independent of and acquired prior to exclusion.

In summary, to examine if, when, and how children revise a theory about how levers work, the present investigation assessed age-related differences in children's judgments of the influence of causal and noncausal variables on the behavior of balance scales (Study 1) and other levers (Study 2). It was predicted that there would be broad, synchronous, and coherent differences with

[2] For all practical purposes, position and orientation are noncausal balance scale variables although, because the centers of gravity are different, a very sensitive balance scale would not perfectly balance when either of these variables is varied on the arms of balance scales. The description of position and orientation as noncausal variables is appropriate given that the manipulation of these features on the arms would so minimally influence the behavior of even a very sensitive balance scale, that the only correct response would be that such balance scales "stayed level" as opposed to "falls to one side or another."

age in children's tendency to make correct judgments not only identifying causal variables which influence the behavior of levers but also dismissing noncausal variables which have no such influence. However, since each alternative seemed plausible, no predictions were made about the nature of children's incorrect judgments and whether or not their judgments about causal and noncausal variables would correlate with each other.

Study 1

A major motivation for specifically studying the role of theory revision in the development of reasoning about balance scales was to understand why 5- and 6-year-olds tend to base their judgments about the behavior of balance scales on the number but not the distances of equally weighted objects on the arms. Such judgments are often explained by young children's lack of a general cognitive capacity to attend to more than one dimension or variable at a time (Inhelder & Piaget, 1958). If children only attended to or "centrated" on the number of objects on balance scales to the exclusion of all other features, then they should dismiss the influence of noncausal variables. Alternatively, if young children's judgments reflect a misunderstanding of how levers work, then they should fail to dismiss the influence of noncausal variables, perhaps even judging that noncausal variables have a causal influence on the behavior of balance scales.

To test whether or not there are age-related differences in children's judgments about the influence of noncausal variables on balance scales, just as there are for causal variables, kindergarten to fifth-grade students judged the behavior of 24 balance scales (see Fig. 1). Children judged the behavior of eight balance scales, the arms of which varied in the causal features of number of equally weighted objects (four Number items) or distance from the fulcrum (four Distance items). Based on previous research, the majority of children in each age group were expected to make correct judgments on Number items, although only a majority of older children were expected make correct judgments on Distance items.

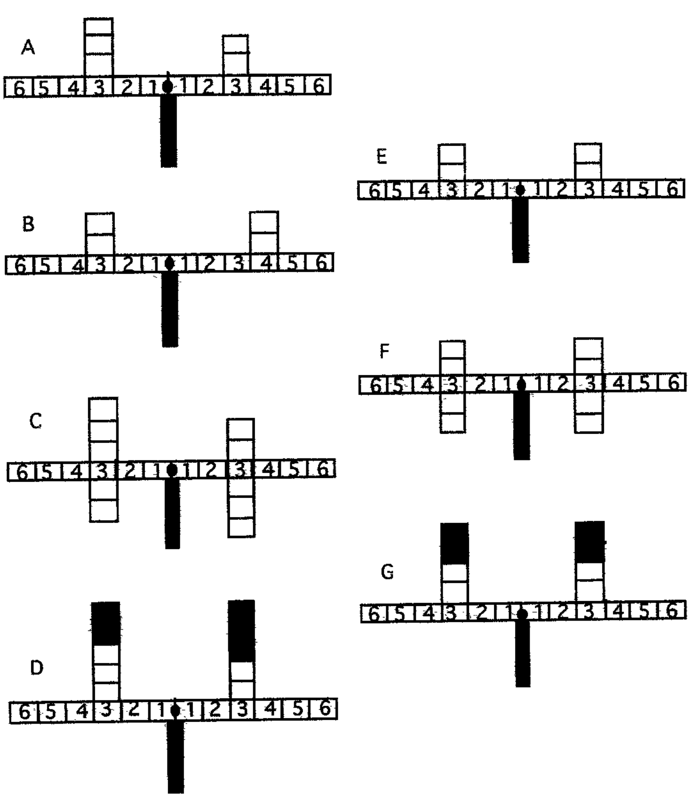

FIG. 1.—Examples of Weight (*A*), Distance (*B*), Position (*C*), Color (*D*), Traditional Balance (*E*), and Novel Balance (*F* and *G*) problems in Study 1

The other 16 balance scales had the same number of blocks placed equidistant from the fulcrum on each arm, so all of them balanced. However, the arms of eight of these balance scales additionally varied in a noncausal variable (see Fig. 1). Four balance scale items varied in the distribution of blocks above and below the arms on each arm (Position items), and the arms of another four varied in the distribution of black-and-white blocks on each arm (Color items). The arms of the final eight balance scales had symmetrical arrangements of blocks on the arms (see Fig. 1). Four Traditional Balance items had white blocks arranged symmetrically on the arms of each balance scale. The blocks on the four Novel Balance items were arranged with the same number of black-and-white blocks on each arm (two items) or with the same number of white blocks placed above and underneath each arm (two items).

Traditional and Novel Balance items have symmetrical arrangements of objects on their arms, and even young children have been shown to correctly judge that balance scales balance when there is such a symmetry (Inhelder & Piaget, 1958; Karmiloff-Smith & Inhelder, 1974; Metz, 1993; Siegler, 1976). Thus, it was predicted that a majority of children in each age group would make correct "balance" judgments on Traditional and Novel Balance items. However, the same "balance" judgment was expected to be made less often on the Color and Position items, whose arms vary in a noncausal feature and so were asymmetrical.

To provide a means of assessing how accurately the features were encoded, children reconstructed each balance scale prior to judging its behavior. The reconstruction task required children to copy the exact arrangement of blocks on each balance scale. In previous studies, Siegler (1976) found age-related decreases in children's reconstruction error rate and a tendency for younger children to make more errors reconstructing information about the distances of objects from the fulcrum than about their number. However, in the present study, children were given the reconstruction task under optimal conditions in order to minimize reconstruction errors. Children were given detailed instructions, practice tasks with feedback, and unlimited time to reconstruct each target balance scale. Siegler (1976) and Siegler and Long (cited in Siegler & Klahr, 1982) found that under such optimal conditions, 5-year-olds made relatively few reconstruction errors and had an equal distribution of number and distance errors, although their judgments about the variables were unaffected.

Of particular interest in the present research was whether or not young children's reconstruction performance under optimal conditions would be the same for all variables. Siegler (1976) listed a number of reasons why children may unsuccessfully reconstruct a balance scale under optimal conditions. None of these reasons for reconstruction errors (e.g., lack of attentional capacity, counting skills, etc.) suggests that children would make more errors reconstructing causal than noncausal features (or vice versa). However, because of their irrelevance, children may encode noncausal features of blocks less carefully than causal ones and so make *more* errors reconstructing the latter than the former features. Alternatively, perhaps because blocks on balance scales are neither typically black-and-white nor typically placed above and below arms, young children may attend more carefully to noncausal features than causal ones and so make *fewer* errors reconstructing the latter than the former features.

Method

Subjects.—Fifteen kindergarten and first-grade students (M = 6 years, 4 months, SD = 6.97 months), 16 second- and third-grade students (M = 8 years, 6 months, SD = 5.94 months), and 15 fourth- and fifth-grade students (M = 10 years, 3 months, SD = 6.31), sampled from a small private Catholic school in a working-class section of New York City were subjects in the experiment. The groups are labeled by the mean age of children in them: 6-, 8-, and 10-year-olds. All children spoke English fluently, and there was an approximately equal number of males and females in each group. One child in kindergarten failed the pretest and was eliminated from the analysis.

Materials.—Two 4 × 4 × 10-inch wooden stands were each glued to a 10 × 10 × ½-inch wooden base. Each stand had two ½-inch eye-hooks screwed onto its top which acted as a fulcrum when an arm, made of a 2 × 2 × 24-inch wooden stick with a ¼-inch hole drilled through the center, was placed between the eye-hooks and a small dowel was inserted through the holes. An arm had 12 alternating black-and-white 2-inch wide sections with a different colored dot placed in the middle of each section.

The same sequence of colored dots was placed on each side of the arm.

One of the two stands was used for the reconstruction task. An arm was glued to the stand and a strip of 3M magnetic tape was attached to the top and the bottom of the arm. Additionally 1½-inch pieces of the magnetic tape were attached to the top and the bottom of 32 2 × 2 × 4-inch wood blocks, half of which were painted white; the other half were painted black. The magnets permitted easy placement of the blocks on top of and underneath each arm on the reconstruction task.

The other stand was used to present the 24 balance scales during the experimental session, four practice items, and three pretest items. Twenty-four arms were prepared by gluing black-and-white blocks on the arms in order to present each item during the experimental session, whereas the twenty-fifth arm was used for the pretest and practice items. The 24 experimental balance scale items comprised six item types (Traditional and Novel Balance items, and variation in Number, Distance, Color, and Position items), each with four problems. The Number, Distance, Color, and Position items had two pairs of problems which were mirror images of each other. For example, one Number problem had four blocks placed on the left and three blocks on the right arm of the balance and another Number item had three blocks placed on the left and four blocks on the right arm of the balance scale, with all blocks placed on the third section. Only the feature of one block differed on the arms of balance scales when color and position were varied, which corresponds to the one unit of difference on the arms of balance scales when number and distance were varied. This equalized the amount of variation when a feature differed on the arms of balance scales.

Procedure.—Each subject was interviewed in a quiet room in the school by one experimenter who initially familiarized the child with the apparatus by naming the parts of the balance scale, pointing out that each side of the arm had six sections (denoted by a colored dot), and giving two blocks to the child and having her report that they were equal in weight (the child was then assured that all the other blocks also weighed the same).

As a pretest, children predicted the behavior of the balance scale with one block placed (in random order) on an arm and at the fulcrum. Children received feedback for each pretest judgment and, if an error was made, the item was presented again in a slightly different form. All children made correct pretest predictions on the first or second presentation of each pretest item.[3] After the pretest, each child was presented with the reconstruction balance scale and told:

> The game we are going to play is called the copy game. It's called the copy game because the goal of the game is to make your balance scale look just like my balance scale. This one is your balance scale (reconstruction balance). This one is my balance scale (experimental balance). Both have the same number of sections. See, there are six sections on this side of my scale and six sections on this side of your scale. Also, there are six sections on this side of my scale and six sections on this side of your scale. In this game you put these blocks on your scale so that it looks just like mine.

Number, position, distance, and color practice items, that were unlike those used in the experimental session, were presented to each subject in a random order. Children were told:

> Remember, try to make the two balance scales look the same. To do that you should look very closely at my balance scale. Make sure that your balance scale has the same number of blocks that are on my scale. Make sure that the same color of blocks are on your scale that are on my scale. Make sure that the blocks on your scale are placed on the same colored dot just like on my scale. Also make sure that the blocks on your scale are placed above or under the arm just like on my scale.

During the practice trials, children received feedback when they made a reconstruction error and, upon completion of a practice reconstruction task, they were asked the judgment question: "If I put the stick through the holes, will the arm stay level, will this side go down, or will the other side go down? Why?" Children's incorrect judg-

[3] It could be argued that the feedback presented for the pretest items may have inflated children's frequency of correct judgments on Number or Traditional Balance items. Such concerns are misdirected since children in the present study and in Siegler (1976), who used no such pretest, made correct judgments at approximately the same rate on these items. Indeed, Siegler (1976, Study 1) found that a full fledged training program (as opposed to a three-item pretest) had no effect on children's balance scale judgment performance unless they had previously been given training to encode the relevant information.

ments on the practice items were *not* corrected.

During the experimental session, the experimenter recorded any errors in subjects' reconstructions. Upon announcing that she had finished reconstructing a balance scale, each child was asked to make a prediction regarding the behavior of the target balance scale and to justify the prediction. In a few cases, a child corrected his or her reconstruction while making a prediction. The presentation of the 24 balance scales was randomized once into six blocks of four problems. These blocks were always presented in the same order, although the sequence of presenting the four problems within each block was randomized for each child. The procedure was completed in one session of about 30 min for all but one child who completed it in two sessions.

Results

Children's judgments about each of the 24 balance scales were coded as correct (scored as 1) or incorrect (scored as 0), summed over item type and subjected to a 3 (age group) × 6 (item type) mixed-model ANOVA corrected for unequal cell sizes (Levine, 1991). There was a main effect of age group, $F(1, 45) = 7.26$, $p < .01$. A Newman-Keuls ($p < .05$) analysis of children's total frequency of correct judgments (out of 24 items) revealed that 6-year-olds made significantly fewer correct judgments overall ($M = 14.60$, $SD = 5.05$) than the 8- ($M = 18.63$, $SD = 3.85$) and 10-year-olds ($M = 19.60$, $SD = 2.90$), who did not differ significantly from each other.

There was also a significant main effect of item type, $F(5, 215) = 22.16$, $p < .001$. A series of t tests ($p < .01$) revealed that more correct judgments were made on both Balance items (Traditional $M = 3.72$, $SD = .89$ and Novel $M = 3.59$, $SD = .93$) than on Color items ($M = 2.96$, $SD = 1.41$), which, in turn, had a significantly higher mean than Distance ($M = 1.61$, $SD = 1.69$) and Position ($M = 2.33$, $SD = 1.63$) items, whose means did not differ significantly from each other. The mean frequency of correct judgments on Number ($M = 3.44$, $SD = 1.07$) items was significantly higher than the means of Position and Distance items, although not significantly different from the other means.

To test for predicted age-related differences in correct judgments, the data were reanalyzed qualitatively, on the basis of the number of children in each age group making at least three out of four correct judgments on each item type[4] (see Table 1). As predicted from previous balance scale research, a large majority of children in each age group made correct judgments on Traditional Balance, Novel Balance, and Number items. Chi-square analyses revealed age-related differences in the number of children who made at least three correct judgments on Number, $\chi^2(2, N = 46) = 6.09$, $p < .05$, and Novel Balance items, $\chi^2(2, N = 46) = 11.59$, $p < .01$. Although the predicted age-related difference in correct performance on Distance items was confirmed, $\chi^2(2, N = 46) = 6.21$, $p < .05$, the items proved more difficult for this sample than had been expected from previous research. Children's poor overall performance on Distance items may have been due to the objects' varying on the arms by the minimal amount of one unit. Ferretti and Butterfield (1986) found that children made correct predictions on distance items more often as the distances of blocks on the arms of the balance scale were made maximally different.

There was no age-related increase in the number of children making at least three correct judgments on Color or Position items in the age range sampled. However, most 8- and virtually all the 10-year-olds correctly judged that Color had no influence on the behavior of balance scales. The same cannot be said about Position; with approximately half of the children in each age group correctly judging that the position of objects does not influence the behavior of balance scales. In a follow-up study, 14 out 15 university undergraduates, all of whom made correct predictions on all other item types, correctly predicted that variation in the position of blocks on the arms of balance scales had no effect on the behavior of balance scales.[5] Thus, correct judgments on the Number and Color items were made by a large majority of 8-year-olds in the present sample, but correct judgments on Distance

[4] The probability of a child's randomly making exactly three correct judgments on an item type is $(4)(.33^3)(.67) = .099$ and of making exactly four correct judgments is $(.33^4) = .012$. Thus the probability of making three or four correct judgments is $p = .111$.

[5] The University of Saskatchewan undergraduates were given paper and pencil versions of the same Number, Distance, Color, and Position items, but no reconstruction task.

TABLE 1

MEAN FREQUENCY OF CORRECT JUDGMENTS BY ITEM TYPE AND AGE GROUP, STUDY 1

Age Group	n	Traditional Balance	Number	Distance	Novel Balance	Position	Color
		Item Type					
6	15	3.33	2.87	.80	2.93	2.33	2.33
		(87)[a]	(73)	(13)	(67)	(47)	(53)
		(1.40)[b]	(1.60)	(1.32)	(1.38)	(1.63)	(1.71)
8	16	3.81	3.75	2.19	3.88	2.00	3.00
		(94)	(100)	(56)	(100)	(44)	(75)
		(.54)	(.45)	(1.80)	(.34)	(1.63)	(1.37)
10	15	4.00	3.67	1.80	3.93	2.67	3.57
		(100)	(93)	(40)	(100)	(60)	(87)
		(.00)	(.62)	(1.79)	(.26)	(1.68)	(1.13)

NOTE.—Maximum score = 4.
[a] Percentage of children who made three or four correct judgments on each item
[b] Standard deviation of the frequencies of correct judgments.

and Position items were made only by a large majority of subjects older than the oldest age group sampled in the present study.

To test the hypothesis that balance scales would be correctly judged to balance less often when their arms varied or a noncausal variable than when their arms did not vary at all, a priori one-tailed *t* tests within each age group were computed comparing performance on Traditional and Novel Balance items with Color and Position items. The hypothesis was confirmed, with children giving correct "balance" judgments more frequently on Traditional and Novel Balance items than Color and Position items (t values ranged from 1.87 to 4.93 with dfs ranging from 14 to 15, all $ps < .05$). The only exception was the 10-year-olds' mean frequency of correct predictions on Color items, which was not significantly different than their means on Traditional, $t(14) = 1.61$, $p < .10$, or Novel Balance items, $t(14) = 1.38$, $p < .10$.

Error analysis.—Children who made incorrect judgments on each noncausal variable were assessed for whether or not they judged that balance scales always fall to the side with a particular level of a noncausal variable. Such an error pattern reflects a false belief that a noncausal variable has a causal influence. However, rather than reflecting a false belief, a majority of children who failed to dismiss the influence of Color and Position made judgments reflecting confusion or uncertainty about the influence of each noncausal variable on the behavior of balance scales. Of the 23 children who failed to make at least three correct balance judgments on Position items, only five (22%) made judgments on all four problems that balance scales would fall to the side of more blocks below (two children) or above (three children) the arm. The other children made a combination of balance and imbalance judgments which reflected neither a causal nor a noncausal belief. Similarly, of the 13 children who failed to make three or four correct predictions on Color items, two (15%) made judgments on all four problems that balance scales would fall to the side of the arm with more black than white blocks. Again, the other children made a combination of balance and imbalance judgments which reflected neither a causal nor a noncausal belief.

A similar error analysis was also performed on children who made incorrect judgments on causal items. Children who failed to identify the influence of each causal item were assessed for whether or not they always made "balance" judgments. Such a pattern of incorrect judgments for a causal variable reflects a false belief that the variable is noncausal. A majority of children who failed to identify number and distance as causal variables did so by making judgments reflecting a false belief that each variable was noncausal. Of the 29 children who failed to make at least three correct judgments on Distance items, 20 (69%, 10 6-, five 8-, and five 10-year-olds) incorrectly judged every time that balance scales would balance when varying in the distance of objects on the arms. Of the five children who failed to make at least three correct predictions on Number items, three (60%) 6-year-olds made only "balance" predictions. Thus, a majority of children who erred on causal variables made judgments reflecting a false belief whereas only a minority of children made judgments reflecting a false belief for noncausal variables.[6]

Relation between causal and noncausal variables.—Pearson correlation coefficients were computed between children's frequency of correct predictions on each causal (Number and Distance) and noncausal (Color and Position) item type (see Table 2). Of particular interest were the correlation coefficients between pairs of causal and noncausal variables which showed similar developmental functions. Age differences in children's judgments for the Number and Color item types and for the Distance and Position item types showed such synchrony. If age differences in judgments about the synchronous variables occurred en masse, reflecting children's acquisition of a new causal understanding of the domain, there should be a positive correlation between the two pairs of variables. Alternatively, if age differences in children's judgments about each pair of variables occurred in a *piecemeal* fashion, there should be no significant correlation between the variables. The data support the latter interpretation, as no correlation coefficient in the matrix was significant.

To further assess the relation between performance on the causal and noncausal items, the frequency of correct judgments on each noncausal item was examined among children categorized as conforming to patterns of judgments identified by previous research (Siegler, 1976). One group of children was identified who made at least three out of four correct judgments on the Traditional Balance item type, at least three out of four correct judgments on the Number item type, and at least three out of four incorrect "balance" judgments on the Distance item type (21 children). This group made balance scale predictions on the basis of the number of blocks on the arms and not their distances from the fulcrum. Another group of children was identified who made at least three out of four correct judgments on the Traditional Balance, Number, and Distance Item Types (15 children).[7] These children made balance scale judgments on the basis of the number and distance of blocks, although some children may have been additionally able to coordinate the causal variables. Seventy-eight percent of the children were categorized as conforming to one or the other pattern, which is compatible with other research (Ferretti & Butterfield, 1986; Ferretti et al., 1985; Siegler, 1976). These two groups of children were not significantly different in age, mean frequency of correct judgments on Color or Position items, or the percentage of children making at least three correct judgments on the noncausal items (Table 3).

Reconstruction errors.—Each reconstructed balance scale was coded as unsuccessfully (scored as 1) or successfully (scored as 0) completed. The mean number of balance scales that children unsuccessfully reconstructed was significantly related to age group in a one-way ANOVA, $F(2, 42) = 3.64$, $p < .05$ (see Table 4). Post hoc analysis revealed that 6-year-olds unsuccessfully re-

[6] Two points need to be raised about the error analysis. First, because subjects' patterns of errors were assessed only if they failed to make correct predictions (e.g., they made at least two incorrect judgments on an item type), a definition of a "false belief" pattern was used which was more conservative than the definition of a "correct" pattern. To be categorized as making errors reflecting a false belief, subjects had to show a false belief pattern on *all* items of an item type, whereas to be categorized as making correct judgments, three out of four items had to be judged correctly. Second, no assessment was presented of the likelihood of a child randomly responding and generating a "false belief" pattern for three reasons. One, it was unclear what assumptions about the prior probability ought to be made to account for the fact that error analysis was performed on children who made at least two judgment errors for a given item type. Two, it was also unclear what assumptions about the prior probabilities were required to account for the different definitions of the "false belief" pattern for causal and noncausal variables. For example, children could have generated a false belief pattern in one of two ways for noncausal variables (e.g., for position, they could have judged that balances fall to the side that has more blocks either above or below the arms) but only in one way for causal variables (e.g., for distance, they had to judge that balance scales balance when the blocks on the arms differ in their distances from the fulcrum). Three, whatever concern there may have been that children generated a "false belief" pattern randomly would be alleviated in Study 2, in which there was a doubling of the number of items for each item type and thus a large reduction in the a priori probability of children generating the "false belief" pattern randomly.

[7] The probability of a child's randomly making judgments conforming to either pattern is $[260(.33^9)(.67^3)] + [66(.33^{10})(.67^2)] + [12(.33^{11})(.67)] + (.33^{12}) = .004$.

TABLE 2

CORRELATIONS BETWEEN CORRECT JUDGMENTS ON EACH CAUSAL AND NONCAUSAL ITEM TYPE, STUDIES 1 AND 2

	NONCAUSAL ITEMS			
	Study 1[a]		Study 2[b]	
CAUSAL ITEMS	Color	Position	Color	Orientation
Number	−.16	−.16	.06	.13
Distance	.23	−.03	.03	.15

[a] In Study 1, there were 46 subjects given four problems for each item type. No correlation coefficient was significant ($df = 44$).
[b] In Study 2, there were 95 subjects given eight problems for each item type. No correlation coefficient was significant ($df = 93$).

constructed more balance scales than 8- and 10-year-olds, who did not differ. The features (number, distance, color, and position of blocks) of each balance scale that were unsuccessfully (scored as 1) and successfully (scored as 0) reconstructed were coded and subjected to a 3 (age group) × 4 (feature) mixed-model ANOVA corrected for unequal cell sizes (Levine, 1991). There was a significant main effect of age group, $F(2, 43) = 164.72$, $p < .001$, a significant main effect of feature, $F(3, 129) = 21.46$, $p < .001$, and a significant age group × feature interaction, $F(6, 129) = 3.91$, $p < .001$. Follow-up t tests ($p < .05$) demonstrated that 6- and 8-year-olds made more errors reconstructing number and distance than position and color. In contrast, 10-year-old children made more distance than number, position, and color reconstruction errors (see Table 4).

Discussion

The results of Study 1 demonstrated the breadth of differences in children's reasoning about balance scales. There were age-related increases in correct judgments not only identifying causal variables but also dismissing noncausal ones. The difficulty young children had in dismissing noncausal variables cannot be attributed to a reluctance to make "balance" judgments. Children had no difficulty making such judgments when blocks were arranged symmetrically on Traditional and Novel Balance items. Indeed, they tended to make balance judgments more often when the blocks were arranged symmetrically than when the blocks varied in a noncausal variable. Moreover, children made few errors encoding the color or position of blocks, suggesting that their correct judgments about

TABLE 3

MEAN AGES AND NONCAUSAL ITEM TYPE JUDGMENTS (and Standard Deviations) OF CHILDREN CONFORMING TO PATTERNS OF JUDGMENTS ON TRADITIONAL ITEM TYPES, STUDIES 1 AND 2

JUDGMENTS ON TRADITIONAL ITEM TYPES[a]	n	AGE (in Months)	NONCAUSAL ITEMS TYPES: Color	Position	Orientation
Study 1:					
Number alone	21	102.29 (17.23)[b]	2.81 (71)[c] (1.57)	2.43 (52) (1.63)	. . .
Number and distance	15	106.20 (18.46)	3.40 (87) (1.12)	2.47 (53) (1.60)	. . .
Study 2.					
Number alone	51	90.35 (18.65)	7.76 (94) (.62)	. . .	5.84 (53) (2.63)
Number and distance	22	119.46 (21.10)	8.00 (100) (.00)	. .	7.05 (82) (1.94)

[a] See text for definitions of these patterns for Studies 1 and 2.
[b] Standard deviations.
[c] Percentage of children who made at least three (Study 1) or seven (Study 2) correct judgments on each item type

TABLE 4

MEAN FREQUENCY OF BALANCE SCALES UNSUCCESSFULLY RECONSTRUCTED BY FEATURE AND AGE GROUP, STUDY 1

Age Group	Balance Scales Unsuccessfully Reconstructed	Causal Features		Noncausal Features	
		Number	Distance	Position	Color
6	9.73[a]	5.13	5.20	1.87	2.27
	(5.64)[b]	(4.16)	(3.73)	(1.73)	(1.71)
8	3.06	1.18	1.94	.44	.31
	(1.61)	(1.05)	(1.69)	(.73)	(.60)
10	2.13	.27	1.73	.20	.13
	(1.81)	(.46)	(1.71)	(.41)	(.32)

[a] The sum of the errors for features is greater than the mean number of balance scales unsuccessfully reconstructured because a balance scale may have more than one feature unsuccessfully reconstructed

[b] Standard deviation

balance scales varying in a noncausal variable were due to the recognition that the variable has no influence, rather than a failure to detect the varying feature.

The results of Study 1 also showed that there is synchrony in children's reasoning about causal and noncausal influences on balance scales. As mentioned previously, the developmental functions in children's judgments on the Number and Color items were similar, with a majority of 6- and virtually all 8- and 10-year-olds correctly judging their influence. In contrast, approximately 50% of the 10-year-olds and almost all the college students correctly judged the influence of distance and orientation. The difference in children's judgments of the influence of number and color compared to distance and position was not due to the causal status of the variables since a causal and noncausal variable constituted the pair that children had more or less difficulty judging. Moreover, the difference in children's judgments about the influence of each pair of variables was unrelated to how well the features were encoded since children tended to more accurately reconstruct one of the features in each pair of variables and less accurately reconstruct the other feature.

We propose that the difference in the development of children's judgments about the influence of number and color, on the one hand, and distance and orientation, on the other hand, was related to the nature of the feature varied for each pair of variables. Number and color are physical features that are associated with the kinds of blocks that were placed on the balance scale, whereas distance and orientation are spatial features that are associated with how and where the blocks were placed on the balance scale. Children correctly assessed the causal or noncausal influence of variables which vary in a physical feature long before doing the same for variables which vary in a spatial feature. Thus, the judgments of 6-year-olds in the present study confirm Metz's (1993) observation that preschool-age children work out what it is about the physical features of blocks (e.g., their weight, size, color, or number) which can balance other ones on a balance scale. However, the same 6-year-olds appear to have little appreciation of the role of spatial features on the behavior of balance scales, tending to make judgments either incorrectly denying or being confused or uncertain about their influence. Thus, the age differences in children's reasoning about balance scales are not only broad and synchronous but also coherent since the differences reflect an expansion of children's understanding of the kinds of features (from physical to both physical and spatial) which influence the behavior of balance scales.

The present data suggest that young children's understanding of why efforts applied to levers balance loads is based on comparing the magnitude of the effort and the load (i.e., the number of equally weighted objects) without much appreciation of the influence of spatial features. As such, the results confirm previous claims that young children's balance scale judgments are based solely on the number or weight of objects on the arms. However, the results disconfirm claims that children make such judgments because of a lack of general cognitive capacity to consider more than one dimension or variable. If children's cognitive capacity was such that they focused solely on the number of objects on the arms of balance scales, they should have made judgments dismissing all other variables as noncausal. This was not the case, particularly for Position items: Many young children neither predicted that position of blocks had a causal influence on balance scales nor dismissed it as noncausal.

Three sources of evidence suggest that separate processes underlie age differences in children's judgments about causal and noncausal variables. First, children's frequencies of correct judgments on causal and noncausal variables were uncorrelated. Moreover, there was no corresponding increase in correct judgments regarding noncausal variables among children grouped according to their developmental level in judging the influence of causal variables. Second, a majority of children who made incorrect judgments on noncausal items did so inconsistently, reflecting a confusion or uncertainty about the influence of the variables. In contrast, a majority of children who made incorrect predictions on causal items did so consistently, reflecting the false belief that the variable had no causal influence. Third, children made fewer errors reconstructing noncausal than causal features of balance scales, suggesting that noncausal features were more carefully attended to than causal ones.

In summary, the data show that there are broad, synchronous, and coherent but unrelated age differences in children's reasoning about causal and noncausal influences on balance scales. Such data are consistent with the claim that the age differences in children's causal and noncausal judgments involve a piecemeal revision of a theory of how levers work. However, there are important limitations in Study 1, including perhaps most significantly a lack of evidence of the generalizability of children's causal reasoning about balance scales to other types of levers. The proposal that children initially lack an appreciation of the influence of spatial features on balance scales should apply to other types of levers as well. Moreover, aspects of the procedure in Study 1 may have adversely influenced children's judgments. For example, children may have made fewer correct judgments about noncausal items than they were capable of because of the presence of Traditional and Novel Balance items. Perhaps if no balance items were included, children would be less sensitive to the lack of symmetry on the Position and Color items and make correct balance predictions on these items more frequently. Finally, the probability of children randomly making at least three correct predictions on an item type was .11 (see n. 4 above) and since 46 subjects were analyzed in this manner on six different item types, a total of 30 categorizations may have been unreliable. The reliability of the results of Study 1 would be improved if there were more problems for each item type.

Study 2

The reliability and generalizability of the findings of Study 1 were assessed in Study 2 by examining children's judgments regarding causal and noncausal influences on balance scales, teeter-totters, arms, and levers. Each of these tasks involves rotational motion around a fulcrum so the number of equally weighted objects and their distances from the fulcrum has a causal influence. The *balance scale* task was much like the task used in Study 1, except that the balance scales were drawn rather than being physically presented. In the *teeter-totter* task, children were presented with drawings that looked like balance scales but with people and not blocks drawn on the apparatus. There were handles on the arms of the teeter-totter which marked off distances from the fulcrum. In the *levers* task, subjects were presented with a drawing of two levers, each of which had a number of blocks placed at various distances from the fulcrum. The *arms* task involved presenting children with a drawing of a person with outstretched arms on which were drawn a number of objects. Distances of the objects from the person's shoulders were marked by stripes on the shirt sleeves.

Children's judgments about balance scales have previously been compared to their judgments on other physics tasks (Ferretti & Butterfield, 1986; Ferretti et al., 1985; Inhelder & Piaget, 1958; Siegler, 1981; Zelazo & Shultz, 1989). There are often weak relations in performance across the tasks, although younger children's judgments often show less of a task effect than do those of older children (Siegler, 1981). This has been explained by young children's lacking knowledge about the task domain and so using very simple judgment strategies on both the balance scale and comparison tasks. However, the comparison tasks used in the previous research were often unrelated to the domain of levers or even simple machines. Because the comparison tasks in the present study are all types of levers, we expected a minimal influence of task on children's prediction performance.

Method

Subjects.—Ninety-five children from elementary schools in a medium-sized western Canadian city were the subjects in the study. There were 12 kindergarten students (M = 5 years, 5 months, SD = 6.11 months), 23 first graders (M = 6 years, 4 months, SD = 6.04 months), 12 second graders (M = 7 years, 5 months, SD = 6.27 months), 18 third graders (M = 8 years, 4 months, SD = 7.39 months), 18 fourth graders (M = 9 years, 3 months, SD = 6.02 months), and 12 sixth and seventh graders (M = 11 years, 8 months, SD = 9.05 months). The groups are labeled by the mean age of children in them: 5-, 6-, 7-, 8-, 9-, and 11-year-olds. Sex was approximately equally distributed in each of the groups. All children spoke English fluently, and no subject failed the pretest.

Materials and procedure.—Children were interviewed individually in a private room in the school and initially pretested on simple versions of the balance scale, teeter-totter, levers, and arms tasks. The pretest not only assessed children's predictions of the behavior of a working model of each task with a single object placed on one arm but also the correspondence between a working model and a drawing of each task. For example, a 3-cm × 14-cm Fisher-Price toy teeter-totter and a 5-cm toy doll were used to pretest children's judgments about teeter-totters. Children were told:

> This is a toy teeter-totter. Have you ever played on a real teeter-totter? Now, let's say I put this little boy on this side of the teeter-totter. Will the teeter-totter go down on this side or the other side, or will the teeter-totter stay level?

After making a judgment, children were shown that the teeter-totter falls to the side on which the doll is placed. If a child's judgment was incorrect, the procedure was repeated until a correct judgment was made. Then the child was asked to judge the behavior of the teeter-totter with the doll placed on the other arm. Once correct predictions were made regarding the behavior of the actual working model of the teeter-totter, children were shown a black-and-white line drawing and told:

> Here is a picture of a teeter-totter, like the one we were playing with. Let's pretend that this is just like a real teeter-totter. Now, when a person gets on this side of the teeter-totter, what will happen? Will the teeter-totter go down on this side or the other side, or will the teeter-totter stay level?

If a correct judgment was made, the child was asked the same question for a person drawn on the other arm. If the child made an incorrect judgment regarding a drawing of a pretest item (which was infrequent), the same question was repeated using both the model and the drawing. The same procedure was used with working models and drawings of a balance scale, levers, and arms.

After the pretest, 32 items were presented to each child in a random order. The items were presented on eight 8.5 × 11-inch black-and-white line drawings of each of four different tasks. For the balance scales, arms, and levers, black or white blocks were drawn as objects, whereas for the teeter-totters, people dressed in black or white clothing were drawn as the objects.

The Distance, Number, and Color items used in Study 1 were again used in Study 2. Instead of varying the position of objects above or below the arms of each task, we varied whether the objects were in a standard horizontal or transformed vertical orientation (Orientation items). Orientation was manipulated because it was difficult to realistically vary objects' position on the teeter-totter task. As it was, varying orientation on the arms of the teeter-totter in the manner that it was varied on other tasks involved depicting people as lying (horizontal) or standing (vertical) rather than in the more usual sitting position. Traditional and Novel Balance items were not used in Study 2 as they were in Study 1. With these item types removed, half of all the items in Study 2 had "balance" as the correct judgment.

Each task included two problems for each item type which were mirror images of each other. Summing over tasks, there were eight problems for each item type, which doubles the number of problems per item type from Study 1. Once again a standard of one unit of difference was established for levers varying in the Number, Color, Orientation, or Distance of objects from the fulcrum of objects. Figure 2 presents a Number item for each task.

Prior to the presentation of the tasks, subjects were shown two identical black-and-white blocks and told.

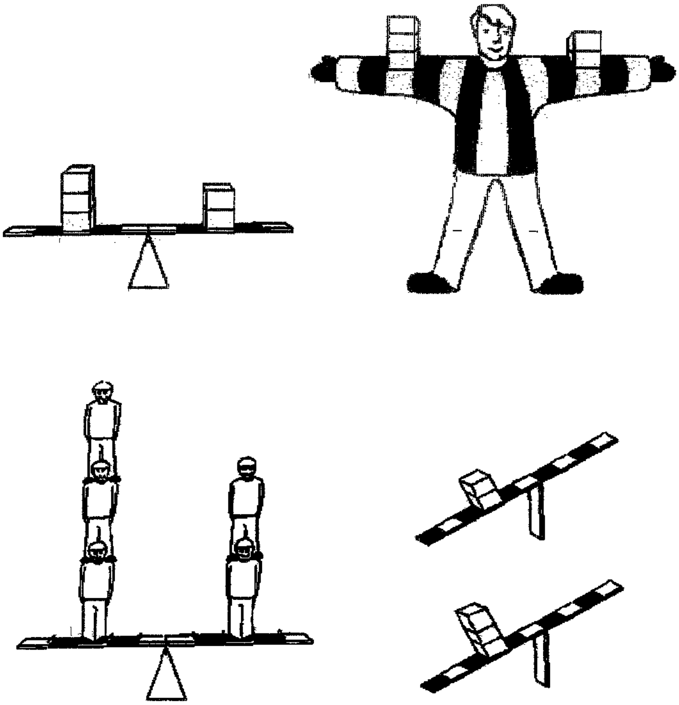

FIG. 2.—Examples of number items for the balance scale, teeter-totter, arms, and lever tasks in Study 2.

> Before we start, look at these two blocks. One is black and one is white, but they are the same size and same weight. Do they feel the same to you? [If a child said no, he or she was told that the blocks were weighed before and that they really were the same]. The blocks in the pictures are like these blocks. Some are black and some are white but they all weigh the same. It is the same for people. Some of the people have white clothes on and some have black clothes on, but they all weigh the same, just like the blocks. Also sometimes the blocks and the people in the pictures are lying down and sometimes they are standing up.

When presented with a balance scale task, children were asked, "Will the balance scale go down on this side or the other side, or will the balance scale stay level?" When presented with a teeter-totter task, children were asked, "Will the teeter-totter go down on this side or the other side, or will the teeter-totter stay level?" When presented with a levers task, children were asked, "Will it be harder to push down on this lever or the other lever, or will the levers be equally hard to push down on?" When presented with an arms task, children were asked, "Will it be harder to hold up this arm or the other arm, or will the arms be equally hard to hold up?" Thus, despite presenting each task as line drawings, which highlights their similarities, the questions posed to children were different for the arms, levers, and the two balance (balance scale and teeter-totter) tasks, which downplays their similarities.

Results

Children's judgments on each of the 32 items were coded as correct (scored as 1) or incorrect (scored as 0) and subjected to a 6 (age group) × 4 (task: balance scale, teeter-totter, arms, and lever) × 4 (item type: Number, Distance, Orientation, and Color) mixed-model ANOVA corrected for unequal cell sizes. There was a significant main effect of age group, $F(5, 89) = 20.66$, $p < .001$. Newman-Keuls ($p < .05$) post hoc analysis revealed that the mean frequency of correct judgments (out of a maximum of 32) of 5- ($M = 18.5$, SD = 4.82) and 6-year-olds ($M = 19.91$, SD = 3.84) was significantly lower than that of 7- ($M = 24.25$, SD = 2.53) and 8-year-olds ($M = 23.78$, SD = 3.92), whose mean frequency of correct performance was significantly lower than that of the 9- ($M = 27.83$, SD = 3.67) and 11-year-olds ($M = 29.83$, SD = 2.73).

There was a significant main effect of item type, $F(3, 267) = 155.70$, $p < .001$, and

TABLE 5

MEAN FREQUENCY OF CORRECT JUDGMENTS BY PROBLEM TYPE AND AGE GROUP, STUDY 2

AGE GROUP	n	ITEM TYPE: Number	Distance	Orientation	Color
5	12	6.58 (66)[a] (2.11)[b]	1.33 (00) (1.78)	3.92 (17) (2.92)	6 67 (67) (2.27)
6	23	7.44 (91) (1.20)	1.30 (04) (2.01)	3.83 (13) (2.44)	7.35 (87) (1.78)
7	12	7.92 (100) (.23)	1.25 (08) (2.34)	7.17 (67) (1.12)	7.92 (100) (.29)
8	18	8.00 (100) (.00)	1.33 (11) (2.68)	6.61 (72) (2.30)	7.83 (94) (.51)
9	18	7.89 (100) (.32)	4.56 (50) (3.75)	7.39 (89) (1.85)	8.00 (100) (.00)
11	12	8.00 (100) (.00)	6.33 (75) (2.67)	7.50 (92) (1.17)	8.00 (100) (.00)

NOTE.—Maximum score = 8.
[a] Percentage of subjects who made correct judgments on at least seven items.
[b] Standard deviation.

a significant age group × item type interaction effect, $F(15, 267) = 5.22$, $p < .001$. Table 5 presents the frequency of correct predictions for each age group for each item type. There was a significant age-related increase in mean frequency of correct predictions on each item type: Number, $F(5, 89) = 4.33$, $p < .01$, Distance, $F(5, 89) = 9.85$, $p < .0001$, Color, $F(5, 89) = 2.55$, $p < .05$, Orientation, $F(5, 89) = 10.81$, $p < .0001$. The age difference in the frequency of correct judgments on each item type concurs with the qualitative analysis of age differences in the frequency of children making at least seven correct predictions on each item type:[8] Number, $\chi^2(5, N = 95) = 19.07$, $p < .01$, Color, $\chi^2(5, N = 95) = 14.35$, $p < .05$, Distance, $\chi^2(5, N = 95) = 36.55$, $p < .0001$, and Orientation, $\chi^2(5, N = 95) = 41.29$, $p < .0001$. As can be seen in Table 5, children correctly judged the influence of Number and Color at an earlier age than they correctly judged the influence of Distance and Orientation, with Orientation being judged correctly prior to Distance.

There was a significant main effect of task, $F(3, 267) = 7.51$, $p < .001$, with follow-up t tests revealing that children made fewer correct judgments about teeter-totters ($M = 5.65$ out of a total of 8, SD = 1.66) than balance scales ($M = 6.10$, SD = 1.44), arms ($M = 6.10$, SD = 1.36), and levers ($M = 5.99$, SD = 1.36), which did not differ from each other (all significant t values $p < .01$).

There was also a significant two-way interaction effect between task and item type, $F(9, 801) = 7.62$, $p < .001$, and a significant three-way interaction effect between age group, task, and item type, $F(45, 801) = 1.66$, $p < .01$. To examine these task-related effects, four separate 4 (task) × 6 (age group) mixed-model ANOVAs, corrected for unequal cell sizes, were performed on children's frequency of correct judgments for each item type. Only Number showed both a significant task, $F(3, 267) = 3.33$, $p < .05$, and a significant age group × task interaction effect, $F(15, 267) = 4.16$, $p < .001$, due to there being age-related increases in mean frequency of correct judgments only on balance scale, $F(5, 89) = 6.93$, $p < .0001$, and teeter-totter, $F(5, 89) = 4.98$, $p < .001$, tasks but not levers and arms tasks.

There was a significant main effect of task for Distance, $F(3, 267) = 3.63$, $p < .05$, and for Orientation, $F(3, 267) = 14.32$, $p < .001$, item types. To analyze the task effect for these item types, t tests were run with a conservative alpha ($p < .01$). The mean number of correct Orientation judgments

[8] The probability of a child's randomly making exactly seven correct judgments on an item type is $(8)(.33^7)(.67) = .002$ and of making exactly eight correct judgments is $(.33^8) = .0002$. Thus the probability of making seven or eight correct judgments by chance is .0022.

(out of two) on the teeter-totter task (M = 1.21, SD = .90) was significantly lower than those on the balance scale (M = 1.56, SD = .77), levers (M = 1.62, SD = .67), and arms (M = 1.54, SD = .77) tasks, which did not differ from each other. With regard to the effect of task on the mean number of correct Distance judgments, no comparison was significant with alpha set conservatively.

The task effects can be characterized as largely due to children performing more poorly on the teeter-totter task than other tasks, particularly on Orientation and (for younger children) Number items. Children's poor performance on the teeter-totter task was probably due to people being depicted on the teeter-totter in a manner which violated how children actually use the apparatus. Children typically sit (rather than stand or lie) on teeter-totters, and multiple children on one arm usually sit in front of one another (rather than standing or lying on top of each other). However, children's frequency of correct judgments on the tasks was positively correlated, independently of age (see Table 6). The significant partial correlation coefficients suggest that, as predicted, children responded similarly across the conceptually related tasks despite being posed different questions.

Error analysis.—As in Study 1, a majority of children who made incorrect judgments on noncausal items did so inconsistently, reflecting uncertainty or confusion about the causal status of the variables. Of the 42 children who failed to make at least seven correct judgments on Orientation items, only six children (17%) showed a pattern of incorrect judgments reflecting the false belief that the orientation of objects influences the behavior of levers. Four children judged that the arms of levers with more vertical than horizontal blocks would be more difficult to raise and more difficult to keep raised, and would fall to that side, and two children judged the reverse. Similarly, only one (13%) out of the eight children who failed to make at least seven correct judgments, falsely believed that color influences the behavior of levers—consistently judging that the arms of levers with more black than white blocks would be more difficult to raise and more difficult to keep raised, and would fall to that side. In contrast, of the 73 children who failed to make at least seven correct predictions on Distance items, 46 children (63%) made judgments reflecting the false belief that objects' distance from the fulcrum has no causal influence on the behavior of levers. These 46 children consistently judged that the arms of levers varying in the distance of objects would nonetheless be equally difficult to raise and to keep raised, and that they would balance. However, none of the six children who failed to make at least seven correct predictions on Number items made incorrect "balance" predictions consistently.

Relation between causal and noncausal variables.—As in Study 1, Pearson correlation coefficients computed between children's frequency of correct judgments on each causal (Number and Distance) and noncausal (Color and Orientation) item were not significant (see Table 2). An analysis of correct judgments regarding the noncausal variables was performed on children grouped according to their judgments regarding causal variables. One group of children was identified who made at least seven correct judgments on Number items and at least seven incorrect "balance" judgments on Distance items (51 children). A second group was identified who made at least seven correct judgments on Number and Distance items (22 children).[9] These two groups of children, which together comprise 77% of the sample, differed significantly in age, $F(1, 71) = 34.57$, $p < .001$, but not in the frequency of correct judgments on Color items or in the percentage of children making at least seven correct predictions on Color items. However, there was a significant difference between the two groups in the percentage of children making at least seven correct predictions on Orientation items, $F(1, 71) = 5.70$, $p < .05$, and a difference between the groups in their mean frequency of correct Orientation judgments which approached significance, $F(1, 71) = 3.71$, $p < .10$. However, when the data were reanalyzed using an ANCOVA procedure with age (in months) as a covariate, there was no significant difference between the groups on Orientation item type.

Discussion

Despite methodological and procedural differences between Studies 1 and 2, the results of Study 2 replicated five central findings of Study 1. First, Study 2 replicates the finding of breadth of age-related differences

[9] The probability of a child's randomly making judgments conforming to either pattern is $[120(.33^{14})(.67^{2})] + [16(.33^{15})(.67)] + (.33^{16}) = .00001$

TABLE 6

PARTIAL CORRELATION COEFFICIENTS FOR CORRECT JUDGMENTS ON EACH PAIR OF TASKS, INDEPENDENT OF AGE, STUDY 2

	Balance Scales	Teeter-Totter	Arms	Levers
Balance scale	. . .	.58*	.64*	.63*
Teeter-totter		. . .	.59*	.58*
Arms			. . .	.74*
Levers				. .

NOTE.—The df = 92 for each partial correlation coefficient
* $p < .001$.

in children's predictions of the influence of causal and noncausal variables on the behavior of balance scales and other types of levers. Second, Study 2 replicates the finding of synchrony in children's judgments of the influence of the number and color of objects placed on the arms of levers on the one hand and the distance and orientation (rather than position) of the objects on the other hand. The majority of 5- and 6-year-olds identified the causal influence of the number of equally weighted objects on the behavior of levers and dismissed the influence of objects' color. However, only a minority of the same children made correct judgments about the influence of the distance of objects from the fulcrum and their orientation. Not until 7 years of age was the influence of orientation correctly dismissed by a majority of children and not until 9 years of age was the influence of distance from the fulcrum correctly identified by a majority of children. Third, the age-related differences in children's reasoning are coherent, reflecting the revision of a theory of the behavior of levers. Children initially work out the influence of physical features of blocks placed on levers (color and number) and only later work out the influence of spatial (orientation and distance) ones, supporting the notion that the age differences in children's reasoning about levers involve the growth of their understanding of features (physical vs. physical and spatial) which influence the behavior of levers. Fourth, the majority of children who made incorrect predictions on noncausal items did so inconsistently, which contrasts with the consistency with which a majority of children made incorrect predictions on Distance items. While a majority of children who made incorrect predictions on Number items in Study 1 did so consistently, this was not the case in Study 2. Fifth, children's frequency of correct judgments on each causal item type was not correlated with their frequency of correct predictions on either noncausal item type. Also, there was no corresponding increase in correct judgments on noncausal variables among children grouped according to judgments regarding causal variables.

General Discussion

The present research assessed children's judgments regarding balance scales and other levers whose arms varied in a causal (the number of equally weighted objects or their distances from the fulcrum) or a noncausal (the color, position, or orientation of objects) feature. The results suggest that there are broad, synchronous, and coherent differences with age in children's judgments about causal and noncausal influences on levers, reflecting the revision of a theory of how levers work. A majority of young children determined which physical features of objects (i.e., the weight or the number of equally weighted objects) influence the behavior of levers and which do not (e.g., the color of objects). This finding supports Metz (1993), who argued that young children work out the physical features of objects which influence the behavior of pan balances. However, only among older children did a majority show a similar determination for spatial features. Young children not only failed to identify the influence of objects' distances from the fulcrum on the behavior of various types of levers but also failed to dismiss the influence of objects' orientation and position. By neither correctly identifying spatial variables as causal nor correctly dismissing them as noncausal, young children showed that they have little appreciation of the role of spatial relations in when, why, and how efforts balance, raise, or overcome loads.

We propose that young children's conception of features which could influence the behavior of levers initially centers on properties of objects placed on their arms and then expands to include the spatial relations between the objects. This pattern of initially reasoning about physical features of objects and later about spatial relations between objects is also found in children's reasoning in other physics domains (Amsel, Savoie, Deak, & Clark, 1991; Reiner et al., 1988; also see Proffitt, Kaiser, & Whelan, 1990, for a related idea). Moreover, the pattern is consistent with previous research on the development of reasoning about balance scales, which has demonstrated that young children initially make balance scale judgments solely on the basis of the number of equally weighted objects and subsequently also take into account the distances of objects from the fulcrum. However, rather than characterizing young children as "centrating" on a single feature (e.g., the number of objects) to the exclusion of other features when making balance scale judgments, the present findings suggest that they have yet to understand the potential relevance of spatial features. That is, children revise, or more aptly expand, their understanding of when, why, and how efforts balance, raise, or overcome loads.

The results further suggest that there was no relation between children's tendencies to identify the causal influence of the number of equally weighted objects and to dismiss the influence of the color of objects, although children learn about the influence (or lack of influence) of each variable at approximately the same age. Similarly, there was no relation between children's tendencies to identify the causal influence of objects' distances from the fulcrum and to dismiss the influence of their position or orientation, although again children learn about the influence (or lack of influence) of these variables at approximately the same age. Moreover, there was no direct relation between children's judgments dismissing the influence of noncausal variables on levers among children who were grouped according to their progression in correctly identifying the influence of causal variables.

It was proposed that independent processes may underlie changes in children's judgments of the influence of causal and noncausal variables if there was no relation between children's judgments about each type of variable. Specifically, it was proposed that different inference skills may underlie children's learning to make correct judgments of the causal (inclusion) or noncausal (exclusion) influences of a variable on an outcome (Inhelder & Piaget, 1958; Kuhn et al., 1988, 1992). As further evidence of the existence of independent processes, consider the different kinds of errors children made when judging causal and noncausal variables and the implications of such differences for the kinds of difficulties they had to overcome in order to make correct judgments. A majority of children in each study who failed to identify the causal influence of distance (and of number in Study 1) made incorrect judgments consistently—they tended to predict that levers varying in these features would balance. Such errors reflect a false belief about the noncausal status of the variables. Moreover, children had more difficulty accurately encoding the causal features, making more errors reconstructing them than noncausal features. Thus, the present research confirms the findings of Metz (1993) and Siegler (1976) in suggesting that to overcome errors judging the influence of causal variables (particularly distance) on lever tasks, children may have to learn to accurately encode causal features and to revise false beliefs regarding their influence.

However, learning to accurately encode and revise false beliefs regarding the influence of causal variables on levers is only half the developmental story. There were also developmental differences in children's ability to correctly dismiss the influence of noncausal variables that were unrelated to differences in their ability to correctly identify causal ones. A majority of children who failed to dismiss the influence of noncausal variables made judgments inconsistently, reflecting confusion or uncertainty about their influence. Furthermore, the reconstruction error data suggest that children carefully attended to the noncausal features, making fewer reconstruction errors for them than the causal features. Perhaps the noncausal features were so salient that children could not ignore their variation, but having no specific knowledge about the influence of variables, they made judgments inconsistently. Children have been shown to have difficulty in a range of other situations where they must ignore salient but irrelevant information (Amsel, Bobadilla, Coch, & Remy, in press—a; Dempster, 1992). For example, children have more difficulty recognizing the undecidability in choosing between two equal alternatives when their choices vary in an irrelevant feature than when they do not vary in this way (Scholnick & Wing, 1988; Somerville, Hadkinson. & Greenberg, 1979). This seems to parallel the present findings that children had more difficulty judging the behavior of levers when their arms varied in an irrelevant feature than when they did not vary at all. Thus, the other half of the developmental story regarding children's reasoning about levers involves their overcoming errors in judging the influence of noncausal variables, perhaps by learning to ignore salient but irrelevant information.

Thus, in addition to suggesting that children form and revise a theory of levers, the present findings suggest that separate processes underlie age differences in children's judgments of causal and noncausal influences on levers. However, two qualifications of this latter conclusion are in order. First, children were not assessed for their ability to coordinate number and distance information on the lever tasks. That is, no task problem varied both the number and distance of objects placed on the arms of levers. We can speculate that perhaps, if such problems were given, there would be a positive correlation between children's correct judgments coordinating number and distance information and dismissing noncausal variables. Children may dismiss all other variables as noncausal except number and distance as they learn to make judgments coordinating these variables. However, about half the children categorized as making judgments on the basis of the number of objects alone, not the distances of objects from the fulcrum, also made judgments correctly dismissing position (Study 1) and orientation (Study 2) as noncausal. Moreover, the percentage of these children who made correct judgments on color items is even higher. Thus, the ability to make judgments coordinating number and distance information may be a sufficient condition for dismissing noncausal variables, but it is not a necessary one.

Second, children's tendency to be confused or uncertain about the influence of noncausal variables on levers tasks does not characterize their judgments about the influence of other noncausal variables on different physics tasks. As previously noted, children tend to make incorrect judgments on gravity-related tasks reflecting a false causal belief that the weight of objects influences the rate at which they fall, roll, and swing. Learning to dismiss the influence of a variable which is judged to be causal is notoriously difficult (Amsel & Brock, in press–b; Kuhn et al., 1998, 1992; Schauble, 1990) and likely to be much more difficult than learning to dismiss a variable which is judged to be neither causal nor noncausal, as was the case with children's incorrect color, orientation, and position judgments. The proposal that children may have more difficulty dismissing a variable erroneously judged to be causal than one judged to be neither causal nor noncausal is supported by the findings that virtually all the adults tested dismissed the influence of noncausal variables on balance scales, but many adults (even physics students) fail to dismiss the influence of weight on gravity-related tasks (Gurstone & White, 1981). However, before any final conclusions are reached, more carefully controlled studies are necessary comparing children's ability to dismiss variables where one variable is initially judged to be causal and the other to be neither causal nor noncausal. This would augment existing research which only examines children's ability to test and revise judgments about variables initially judged to be causal or noncausal (Amsel & Brock, in press–b; Kuhn et al., 1998, 1992; Schauble, 1990).

In summary, the results of the present research demonstrate that children's processes for identifying causal and dismissing noncausal influences on levers are independent of each other, but dependent on the kinds of features (physical or both physical and spatial) which children conceive of as potentially having an influence on the behavior of levers. Such findings may well apply to other physics domains than levers (see Amsel, 1995, and Amsel, Clark, Savoie, Gesell, & Mouring, 1996, for a discussion of the parallels between the growth in children's understanding of levers and gravity in terms of identifying causal and dismissing noncausal variables). More generally, the findings may help to clarify a debate about the process of theory revision in one decidedly "unphysics" domain. According to some theorists, children acquire a representational theory of mind when they learn to encode and identify the influence of an agent's (self or other) mental states (e.g., mistaken belief about true states of the world) on the agent's subsequent behavior (Gopnik & Astington, 1988; Perner, 1991). However, other theorists claim that children understand the causal relation between mental states and subsequent behavior but may be unable to ignore, and hence fail to dismiss, the influence of nonmental states (e.g., true states of the world) on an agent's behavior (Chandler, Fritz. & Hala, 1989; Lewis &

Mitchell, 1994; Russell, Mauthner, Sharpe, & Tidswell, 1991). In support of both positions, the results of the present research suggest that revising a theory may involve both learning to correctly encode and identify causal influences on phenomena (e.g., mental states on behavior) and to ignore and dismiss noncausal influences (e.g., nonmental states on behavior). However, the processes for learning to do one may be independent of those for learning to do the other.

One important implication of the present research is its consideration of ways in which theory revision may proceed in children. The present findings suggest that theory revision may occur in a piecemeal fashion, involving an independence between identifying causal and dismissing noncausal variables. This is not to say that children misunderstand the mutual exclusivity that exists in judging a particular variable as causal or noncausal on a specific occasion. Rather, despite having such an understanding, children may employ different processes for and may be confronted by different problems when learning to correctly identify causal and dismiss noncausal influences on theory-related phenomena. Additional longitudinal research will be necessary in order to fully confirm these claims. Nonetheless, the upshot is that while intuitive theories are integrated networks of beliefs and concepts, such holism may not extend to processes underlying theory revision.

References

Amsel, E. (1995, March). The development of causal attributions in two physics domains In E. Amsel (Chair), *Process and content in the development of scientific reasoning* Symposium conducted at the biennial meeting of the Society for Research in Child Development, Indianapolis.

Amsel, E., Bobadilla, W., Coch, D., & Remy, R. (in press–a). Young children's memory for the true and pretend identities of objects. *Developmental Psychology*.

Amsel, E., & Brock, S. (in press–b). Developmental changes in children's evaluation of evidence. *Cognitive Development*.

Amsel, E., Clark, M., Savoie, D., Gesell, S., & Mouring, F (1996). *Development of children's understanding of freefall* Unpublished manuscript, Weber State University.

Amsel, E., Savoie, D., Deak, G., & Clark, M. (1991). Preschoolers' understanding of gravity. *Proceedings of the thirteenth annual conference of the Cognitive Science Society* (pp. 600–605). Hillsdale, NJ: Erlbaum.

Ausubel, D P, & Schiff, H M. (1954). The effect of incidental and experimentally induced experience in the learning of relevant and irrelevant causal relationships by children. *Journal of Genetic Psychology*, **84**, 109–123

Brewer, W., & Samarapungavan, A. (1991). Children's theories vs. scientific theories: Differences in reasoning or differences in knowledge. In R. R. Hoffman & D. S. Palermo (Eds.), *Cognition and the symbolic processes: Applied and ecological perspectives* (pp 209–232). Hillsdale, NJ: Erlbaum.

Bullock, M (1985). Causal reasoning and developmental change over the preschool years. *Human Development*, **28**, 169–191.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. Friedman (Ed), *The developmental psychology of time* (pp. 209–254). New York. Academic Press

Carey, S (1985a). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (1985b). Are children fundamentally different kinds of thinkers and learners than adults? In S. F. Chapman, J. W. Segal, & R. Glaser (Eds.) *Thinking and learning skills: Research and open questions* (Vol **2**, pp 485–517). Hillsdale, NJ. Erlbaum.

Carey, S. (1991). Knowledge acquisition. Enrichment or conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind. Essays on biology and cognition* (pp. 257–292). Hillsdale, NJ. Erlbaum.

Case, R. (1985) *Intellectual development. Birth to adulthood*. New York: Academic.

Chandler, M., Fritz, A, & Hala, S (1989). Small scale deceit. Deception as a marker of three-, four-, and five-year-olds' theory of mind. *Child Development*, **60**, 1263–1277.

Dempster, F. (1992). The rise and fall of the inhibitory mechanism. Toward a unified theory of cognitive development and aging *Developmental Review*, **12**, 45–75.

Ferretti, R. P, & Butterfield, E. C. (1986) Are children's rule-assessment classifications invariant across instances of problem type? *Child Development*, **57**, 1419–1428

Ferretti, R. P, Butterfield, E. C., Cahn, A., & Kerkeman, D. (1985). The classification of children's knowledge: Development on the balance-scale and incline-plane tasks. *Journal of Experimental Child Psychology*, **39**, 131–160.

Gelman, S A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, **23**, 183–209.

Gopnik, A., & Astington, J. (1988) Children's understanding of representational change and its relation to the understanding of false belief and the appearance/reality distinction. *Child Development*, **59**, 26–37.

Gopnik, A., & Wellman, H. M. (1992). Why children's theory of mind really is a theory. *Mind and Language*, **7**, 145–171.

Gurstone, R., & White, R. (1981). Understanding gravity. *Science Education*, **65**, 291–299.

Howe, C., Tolmie, A., & Rodgers, C. (1992). The acquisition of conceptual knowledge in science by primary school children: Group interaction and the understanding of motion down an incline. *British Journal of Developmental Psychology*, **10**, 113–130.

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

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

Keil, F. (1992). The origins of an autonomous biology. In M. R. Gunnar & M. Maratsos (Eds.), *Modularity and constraints in language and cognition: The Minnesota symposium on child psychology* (Vol. **25**). Hillsdale, NJ. Erlbaum.

Klahr, D., & Siegler, R. S. (1978). The representation of children's knowledge. In H. Reese & L. P. Lipsitt (Eds.), *Advances in child development*. New York: Academic Press.

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

Kuhn, D., Schauble, L., & Mila-Garcia, M. (1992). Cross domain development of scientific reasoning. *Cognition and Instruction*, **4**, 285–327.

Levine, G (1991). *A guide to SPSS for analysis of variance*. Hillsdale, NJ: Erlbaum.

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

Metz, K. (1993). Preschoolers' developing knowledge of the pan balance: From new representations to transformed problem-solving. *Cognition and Instruction*, **11**, 31–93.

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

Proffitt, D. R., Kaiser, M. K., & Whelan, S. M. (1990). Understanding wheel dynamics. *Cognitive Psychology*, **22**, 342–373.

Reiner, M , Chi, M., & Resnick, L. (1988) Naive materialistic belief: An underlying epistemological commitment. *Proceedings of the tenth annual conference of the Cognitive Science Society* (pp. 544–551). Hillsdale, NJ: Erlbaum.

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

Schauble, L. (1990). Belief revision in children. The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, **49**, 31–57.

Scholnick, E. K., & Wing, C. S. (1988). Knowing when you don't know: Developmental and situational considerations. *Developmental Psychology*, **24**, 190–196.

Selman, R., Krupa, M., Stone, C., & Jacquette, D. (1982). Concrete operational thought and the emergence of the concept of unseen force in children's theories of electromagnetism and gravity. *Science Education*, **66**, 181–194.

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, **47**(1, Serial No. 194).

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, **8**, 481–520.

Siegler, R. S. (1978). The origins of scientific reasoning. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 109–151). Hillsdale, NJ: Erlbaum.

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, **46**(2, Serial No. 189).

Siegler, R. S. (1985). Encoding and the development of problem solving. In S. F Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills. Research and open questions* (Vol. **2**, pp. 161–185). Hillsdale, NJ: Erlbaum.

Siegler, R. S., & Klahr, D. (1982). When do children learn: The relation between existing knowledge and the ability to acquire new knowledge. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. **2**, pp. 121–211). Hillsdale, NJ: Erlbaum.

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

Siegler, R. S., & Taraban, R. (1986). Conditions of applicability of a strategy choice model. *Cognitive Development*, **1**, 31–51.

Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition*, **21**, 177–237.

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

Surber, C. F., & Gzesh, S. M. (1984). Reversible operations in the balance scale task. *Journal* *of Experimental Child Psychology*, **38**, 254–274.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, **24**, 535–585.

Wellman, H., & Gelman, S. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, **43**, 337–375.

Wilkening, F., & Anderson, N. H. (1982). Comparison of two rule assessment methodologies for studying knowledge development and cognitive structure. *Psychological Bulletin*, **92**, 215–232.

Wilkening, F., & Anderson, N. H. (1991). Representation and diagnosis of knowledge structures in developmental psychology. In N. H. Anderson (Ed.), *Contributions to information integration theory: Vol. 3. Developmental* (pp. 45–80). Hillsdale, NJ: Erlbaum.

Zelazo, P. D., & Shultz, T. R. (1989). Concepts of potency and resistance in causal judgments. *Child Development*, **60**, 1307–1315.