Cognition, 21 (1985) 177–237

On differentiation: A case study of the development of the concepts of size, weight, and density*

CAROL SMITH
University of Massachusetts/Boston

SUSAN CAREY
Massachusetts Institute of Technology

MARIANNE WISER
Clark University

Abstract

This paper presents a case study of 3- to 9-year-old children's concepts of size, weight, density, matter, and material kind. Our goal was to examine two claims: (1) that individual concepts undergo differentiation during development; and (2) that young children's concepts are embedded in theory-like structures. To make progress on the first issue, we needed to specify in representational terms what an undifferentiated concept is like and in what sense this undifferentiated concept is a parent of the more differentiated concepts. Our strategy was to use a model of conceptual differentiation suggested by the history of science to guide our search for evidence. In this model, undifferentiated concepts, like differentiated concepts, can be analyzed in terms of their component properties, features, or dimensions. The key difference is that an undifferentiated concept unites certain components which will subsequently be analyzed as components of distinct concepts, and that the undifferentiated concept is embedded in a different theoretical structure from the differentiated concepts. In our study, the same group of 78 children (18 3-year-olds, 18 4-year-olds, 18 5-year-olds, 12 6–7-year-olds, and 12 8–9-year-olds) were given a range of tasks probing their understanding of size, weight, and density; a

*This research was funded by NSF Grant #SED-7913278, Joint NIE/NSF Research Program. We thank Renee Baillargeon, Ned Block, Rachel Falmagne, Arthur Millman, Helen Tager-Flusberg and two anonymous reviewers for their very careful reading of the manuscript and their many helpful suggestions for revisions. We also thank the teachers and children at the Cambridge Montessori School, Buckingham, Browne & Nichols School, Corner Co-op Nursery School, and Radcliffe Child Care Center for their participation in the study. David Dickinson, Denise LaBuda, Mara Levin and Phil O'Neil provided valuable research assistance.

Reprint requests should be sent to: Carol Smith, Department of Psychology, University of Massachusetts/Boston, Boston, MA 02125, U.S.A.

0010-0277/85/$18.80

subgroup of these children were given additional tasks probing their concepts of matter and material kind. We found that young children had a theoretical system which included distinct concepts of size, weight, and material kind and were beginning to form generalizations relating these concepts (e.g., size is crudely correlated with weight, steel objects are typically heavy). The core of their weight concept was felt weight, with density absent from their conceptual system; material kinds were defined in terms of properties which characterize large scale chunks of stuff. Slightly older children (5–7-year-olds) had made modifications to their concepts of weight and material kind. At these ages, their concept of weight now contained both the properties heavy *and* heavy for size *(which we take as evidence of their having an undifferentiated weight/density concept) and they were coming to see weight differences as important in distinguishing whether large-scale objects were made of the same kind of stuff. However, the core of their weight concept was still felt weight and material kinds were still defined in terms of properties of large scale objects. Finally, still older children (8–9-year-olds) had a theoretical system in which weight and density were articulated as distinct concepts, material kinds were reconceptualized as the fundamental constituents of objects, and weight was seen as a fundamental property of matter. We conclude that children's concepts of weight and density do differentiate in development and that it does make sense to view children's concepts in the context of theory-like structures.*

Introduction

At the core of cognitive development is knowledge acquisition. The most obvious difference between a 4-year-old child and an adult, for example, is that the adult knows more. And the most striking difference between a novice physicist and an expert physicist is that the expert knows more physics. This truism begins to pose a scientific problem when one notes that knowledge acquisition is not the mere accumulation of new facts—knowledge is restructured in the course of its acquisition.

Recently cognitive psychologists have taken up the challenge of characterizing this restructuring. Chi, Glaser & Reese (1982) argue that novices and experts represent different relations among core concepts and that experts represent superordinate concepts that the novice lacks. Although both types of reorganization undoubtedly occur, even more radical restructuring is possible. The history of science has been marked by theory changes in which the core concepts of the theories are themselves modified. The most common type of conceptual change is probably differentiation, such as Galileo's distinguishing *average velocity* and *instantaneous velocity* (Kuhn, 1977) or Black's differentiating *heat* and *temperature* (Wiser & Carey, 1983). Other kinds of conceptual change are part of theory change as well. Sometimes two concepts are coalesced, as in Galileo's collapsing Aristotle's concepts of natural motion and artificial motion into a single concept of motion. And sometimes concepts originally thought to be simple properties are reanalyzed as relations, as in Newton's treatment of *weight*.

This paper concerns one type of conceptual change—differentiation. Differentiation is a ubiquitous process, invoked in the description of many different kinds of development. In biology, embryological development is seen as involving successive differentiations. And psychologists appeal to differentiation in several contexts. Some psychologists have taken entire representational systems as their unit of analysis (e.g., Werner, 1948). Other units of analysis have been sensory systems (e.g., Bower, 1974), percepts (e.g., Gibson, 1969; Kemler, 1983), and concepts (e.g., Piaget & Inhelder, 1974).

Whatever the unit of analysis, differentiation is the progression from a single parent to two or more descendants. Any account of the process of differentiation must provide answers to two descriptive questions: (1) What is the descent relation—that is, in what sense is the parent a precursor of the descendants? and (2) In what sense is the parent undifferentiated with respect to its descendants? These questions receive straightforward answers in the embryological case. The primary unit of analysis in embryological development is the cell, even though the theory explains organ differentiation as well. This is because the mechanisms of differentiation operate within individual cells—various types of environmental factors (mainly chemical) trigger cell differentiation via genetic mechanisms. Descent is interpreted literally; cells give birth to descendants by division. The way parent cells are undifferentiated with respect to their descendants is also fairly clear. The parent cells are all alike with respect to certain properties, making them a distinctive type of cell. Descendant cells keep some of the properties that were specific to the parent. At the same time, they develop new specialized properties not found in the parent which make them into two or more distinctive kinds of cells.

Unlike biologists, developmental psychologists offer no account of the mechanisms of differentiation and provide less clear answers to the questions of tracing descent and characterizing how the precursor is undifferentiated with respect to its descendants. Consider the case of conceptual differentiation. Little explicit attention has been given to the descent relation; more attention has been given to the issue of characterizing an undifferentiated concept. Undifferentiated concepts are said to be diffuse, syncretic, and holistic, relative to their descendants. These descriptions are used in an attempt to specify how the ancestors contain the descendants. The properties which will differentiate the descendant concepts are not distinct in the parent concept; rather, the parent concept has the potentiality for developing those distinct properties, containing them in a more primitive fused form. So, for example, Piaget and Inhelder (1974) wrote that the child initially has only a diffuse, undifferentiated concept of global quantity, from which he subsequently differentiates the concepts of size and weight:

> ... he arrives at the idea of a global quantity before he can construct such differentiated quantities as weights and volumes ... (p. 4)
>
> ... children at this stage of development are incapable of quantifying weight, volume, and even the apparent quantity of matter and hence grasping their conservation ... they fuse all three into an egocentric and phenomenalistic whole ... (p. 161)
>
> ... our present subjects undoubtedly believe that the heavier substance A must have something 'more' than the lighter substance B, the more being something undifferentiated and global ... (p. 161)

This description of undifferentiated concepts as "diffuse, syncretic wholes" commits the psychologist to the claim that undifferentiated concepts are a different *kind* of concept than differentiated ones—the former are diffuse and syncretic while the latter are discrete and analyzable in terms of components. However, Kuhn (1977) denies that conceptual differentiation in the history of science involves a change in type of concepts. That is, he denies that undifferentiated concepts are "diffuse, syncretic wholes" or are in any sense intrinsically "confused". Rather, they function within their theories just as do all scientific concepts. An undifferentiated concept has components which will become specific to each of its descendants. Yet as parts of the undifferentiated concept, those components make up an articulated, integrated and consistent whole. The lack of distinction between those components makes the concept inadequate in some contexts, but that is true of any concept in any theory. The descendants in turn have united components which may subsequently have to be distinguished to make sense of some further phenomena. Thus, it is the concept-as-applied to the world which leads to confusion; not the the concept itself.

Kuhn's account of conceptual differentiation thus provides an alternative account of the sense in which parent concepts contain descendant concepts. The historical case described by Wiser and Carey (1983) can serve as an example. They studied the historical period between the time of the first systematic use of the thermometer by the members of the Florentine Academy in the 17th century and the time of Joseph Black, who first distinguished *heat* and *temperature* in the 18th century. The undifferentiated thermal concept of the Florentine Academy included distinct components of both Black's *heat* (e.g., heat causes boiling) and also of Black's *temperature* (e.g., felt hotness) (Wiser & Carey, 1983). But if we think of such components of a concept as distinct and already present in the parent concept, we are led into a quandary. Why does it make sense to say there is a lack of differentiation? Why not pick the components as the level of description, in which *heat* and *hotness* were clearly distinct, even if they were not the modern concepts of *heat* and *temperature*? On this alternative analysis, there would be no *differentiation*, only changes of concepts with further gains of knowledge. As Wiser and Carey show (following Kuhn, 1977), the resolution of the paradox is to examine the role of the components in the theory in which they are embedded. The theoretical contexts in which the undifferentiated concept *heat/temperature* played a role called for only one concept, one in which these components played no distinct roles. That is, these components were not appealed to differentially in the explanation of different phenomena, did not figure in separate laws, and were not explicitly related to each other in the theory. In contrast, Black explained some phenomena in terms of heat (e.g., the amount of heat necessary for snow to be changed to water) and some in terms of temperature (e.g., the temperature at which water freezes or boils) and he had laws explicitly relating the two. This is why he is credited with drawing the distinction between heat and temperature. Thus, it is only by analyzing concepts relative to the theories in which they are embedded that we can decide how components are packaged, whether in any given case there is one concept or two. Without such analysis, one cannot know whether conceptual differentiations or coalescences have occurred.

Analysis of concepts in terms of components may seem to be incompatible with an analysis of concepts relative to their theories. The two accounts are incompatible, however, only on extreme versions of each program: that is, an analysis-into-components approach which requires that all components be theory neutral and an analysis-relative-to-theories approach committed to extreme meaning holism, such that all components must change when a theory changes. On the latter account, any theory change necessarily entails conceptual change, and the task of tracing descent of individual concepts from one theory to another is impossible. Many historians and philosophers of science, however, reject such extreme views of theory change (see Suppe, 1974, for an extensive discussion of these issues). Both Kuhn (1982/83) and Wiser and Carey (1983) maintain an analysis of theory change that involves true conceptual change in core concepts, although it allows for tracing descent between them. Because successive theories differ in the domains of phenomena accounted for and in explanatory structure, the core concepts in each are not intertranslatable. Descent can be traced, however, because of several properties of theories that stay fixed through change. First, successive theories can agree on some data, and some phenomena are in the domains of both. Second, some components of core concepts in successive theories remain the same and play similar roles in the explanation of common phenomena.

In this paper, we explore the possibility that conceptual change in childhood, like conceptual change in science, needs to be understood in the context of theory change. Although Piaget and Inhelder (1974) do study the development of the concepts of size, weight, and density in the context of the child's emerging atomistic theories of matter, they do not *use* their characterization of the child's theory at a given stage to resolve issues of whether there is one conceptual unit or two, or to trace lines of descent. Indeed, they explicitly assume that the child only gradually develops the abilities to have true concepts (with the advent of concrete operations) and true theories (with the advent of formal operations). Thus, they implicitly endorse the possibility that early undifferentiated concepts are different in kind from later, differentiated, ones. If this were true, conceptual differentiation in early childhood differs markedly from conceptual differentiation in the history of science. If concepts in early childhood cannot be viewed as analyzable in terms of components or as participating in theories, one will have to resort to other means to trace lines of descent and to describe the undifferentiated state.

Two kinds of questions about conceptual differentiation in childhood are raised by the analogy to conceptual differentiation in the history of science. The first is theoretical: Does it really make sense to consider children's concepts as true concepts embedded in theories? The second is evidential: Given that children do not formulate and record research programs, how are we to know when concepts kept separate by adults are conflated by children?

With respect to the theoretical question, there is a growing consensus that Piaget and Vygotsky were wrong in their claims that young children's concepts differ in kind from adults' concepts. Some of the evidence taken to support the claim presupposes a mistaken view of the nature of adult concepts (Fodor, 1972). Other evidence involves tasks with extraneous sources of difficulty, such as the requirement of sophisticated metaconceptual understanding or the knowledge of particular concepts. When these sources of difficulty are removed, preschool children's concepts are seen to be the same in kind as those of adults (Carey, 1984; Gelman & Baillargeon, 1983; Mandler, 1983; Markman, & Callahan, 1983, for recent reviews of the relevant literature). The issue of whether young children hold theories is more complex and has not been systematically addressed. Nonetheless, there is increasing evidence that young children do possess abilities necessary for theory formation which the earlier developmental literature had considered beyond their grasp (Braine & Rumain, 1983; Carey, 1984). Young children have also been shown to have an interest in explanation, as well as a concept of physical causality (Bullock, Gelman, & Baillargeon, 1982; Shultz, 1982). And many science educators are taking the child as scientist metaphor seriously by assuming that elementary age children have theories for interpreting phenomena which are alternatives to the currently accepted theories in science (see Driver & Erickson, 1983, for a review of such studies). Although there is no doubt that children lack metaconceptual awareness of theory construction and evaluation (Carey 1984; Inhelder & Piaget, 1958), in this paper we take it as a working assumption that even preschool children possess theory-like conceptual structures. Our success or failure in our effort to reconstruct a portion of their physical theory will allow us to evaluate this assumption.

Let us turn now to what constitutes evidence for lack of differentiation. Previous workers have taken certain kinds of behavioral evidence to support a claim of nondifferentiation. For example, some of Gibson's evidence that children do not differentiate *size* and *weight* was that their performance on size and weight seriation tasks was identical. In both cases they simply seriated items according to size. Similarly, some of Piaget's and Inhelder's evidence involved size intrusions on weight judgment tasks. For example, some children thought that popcorn became heavier when it was popped because it got bigger. They also predicted that a wax and clay ball would weigh the same if they were made the same size. However, this behavioral evidence is consistent with several different representational states of affairs. These data could result from the absence of *weight* from the child's conceptual system, rather than from *size/weight* lack of differentiation. The data are also consistent with the child's thinking that relative size is a reliable indicator of weight. Indeed, Piaget and Inhelder talk about children's early size and weight concepts in two different ways. At times they use the language of differentiation and imply the child has one undifferentiated size/weight concept (see earlier quotes). At other times they imply that children have perfectly distinct size and weight concepts which they think are directly proportional:

> ... We see at once that none of these children has the least idea of conservation of weight; they all think that the seed should gain weight with increases in volume because they quite generally believe that the two are proportional ... (p. 121)

In a representational account these two descriptions (that *size* and *weight* are undifferentiated and that *weight* is thought to be proportional to *size*) correspond to two quite different and incompatible conceptual states: one involving a single concept and the other involving two distinct concepts.

The analysis of differentiation we are offering, in terms of changes in the packaging and articulation of conceptual components as a function of theory change, has methodological consequences. To determine whether two components are conflated in one concept or belong to separate concepts, one must consider the theory in which the components function. The components of an undifferentiated concept function as a single, integrated unit within the theory. This means that there should be no distinguishable contexts in which the components are separately and systematically applied. Further, there should be some contexts in which both components are concurrently and unsystematically applied to understanding the same phenomena, leading to what looks like confusions relative to later conceptual states. Finally, the components should never be separately related to other concepts in the theory and the theory should contain no explicit beliefs about the interrelations of these components. Clearly, diagnosing conceptual differentiation in children requires multiple tasks which probe their understanding of a variety of concepts representative of a domain and their understanding of the concepts' interrelations. Multiple tasks are also necessary to distinguish lack of differentiation from concept absence. If one concept is absent in the initial theory, there should be no context in which even a component of that concept is applied.

This paper presents the beginnings of a case study of the development of the concepts of size, weight, and density in childhood. We ask whether *size* and *weight* and whether *weight* and *density* are ever undifferentiated in the course of development, and if so, how such undifferentiated concepts should be described. The theoretical context for *size*, *weight*, and *density* studied by Piaget and Inhelder is an atomistic theory in which schemes of compression and decompression of particles are used to explain density differences between substances. Piaget and Inhelder (1974) claimed that this atomistic conception does not emerge until adolescence, and recent work suggests that several aspects of an atomistic model continue to pose problems for high school students (Novick & Nussbaum, 1981; Pfundt, 1981, 1982). The theoretical context we explore in this paper concerns more basic aspects of an adult's concepts of matter and material kind. We assume that the adult understands matter as the stuff objects are constituted of; a table made out of wood is wood all the way through. The adult knows that not everything is a material object—tables and people are constituted of matter; holes, shadows, and areas are not—and that only material objects have weight. They know that a bigger object is heavier than a smaller object of the same material kind because it contains more matter and because matter has weight.

Finally, they also know that matter comes in different kinds which differ in the weight of any fixed volume. Thus, they can explain the weight of an object in terms of the density of the material it is made of and its size.

Our tasks do not require quantitative understanding of any of the three concepts. That is, children were never required to measure stimuli, nor were they required to know units such as inches, pints, ounces, or ounces per cubic inch. Rather, the tasks required qualitative comparisons along the dimensions of size, weight, and density. The density tasks required a concept of weight relativized to size that is seen as a characteristic of different kinds of stuff. The *density* probed here, then, is *heavy for size* and figures in such generalizations as "steel objects are heavy for their size; objects made of balsa wood are light for their size."

Our study is reported in four sections. The first evaluates the claim that children do not initially differentiate *size* and *weight* and the second evaluates the claim that children do not initially differentiate *weight* and *density*. We devised tasks that require responses based on size, weight, and density. Possible evidence for lack of differentiation is intrusions of one factor into judgments calling for the other (e.g., basing a "heavier" judgment on relative size, or basing a "denser" judgment on absolute weight). In Section 3 we attempt to replicate one of the tasks where Piaget and Inhelder found size intrusions on judgments about relative weights. We explore whether young children have any generalizations relating size and weight, and if so, whether they have a theory driven expectation of a perfect correlation between size and weight or an empirically driven expectation of only a rough correlation. Finally, Section 4 concerns these same children's understanding of matter and material kinds. We probe their understanding of the distinction between material and immaterial objects, and the distinction between objects and the substances from which objects are made.

1. The size and weight tasks

If young children have a single concept, combining the components of size and weight, we should observe intrusions of weight into size judgments and intrusions of size into weight judgments. In earlier work, Piaget and Inhelder (1974) found both kinds of intrusions. However, the tasks they gave young children were complex. In one, where size intruded upon a weight judgment, children were asked to make a clay ball that would weigh the same as a wax ball. Success on this task requires a conceptualization of density. Children may well have distinct size and weight concepts but inappropriately believe that size is a good indicator of weight, even when objects are made of different materials. In another, where weight intruded upon a size judgment, the child was asked to predict how high the water would rise when objects of varying sizes and weights were placed in a container. Success on this task requires that the child knows what physical variables affect water displacement. Children may have distinct size and weight concepts but falsely believe that the weight of objects affects the amount of water displaced. Thus, the Piagetian phenomena can be explained without appealing to *size/weight* lack of differentiation.

If the child truly fails to distinguish *size* from *weight*, size and weight intrusions should persist even on simpler tasks. We devised four such tasks. Two probed for the concepts nonverbally—children were asked to predict which objects would fit into a box and which objects would make a foam rubber bridge collapse. Two probed the child's understanding of specific words, "larger" and "heavier". In each task there were critical and noncritical items. Critical items were objects that were big and light, or small and heavy—that is, items in which size and weight conflicted. Noncritical items were objects that were big and heavy or small and light—that is, items in which size and weight covaried. If children have an undifferentiated size/weight concept, in each task they should make more errors on the critical than noncritical items.

Method

Subjects

The subjects were 78 children—18 3-year-olds, 18 4-year-olds, 18 5-year-olds, 12 6–7-year-olds, and 12 8–9-year-olds. The mean age for each group was 3;2, 4;4, 5;4, 6;11, and 9;3, respectively. All 5–9-year-olds and several 3- and 4-year-olds were recruited from a private day school in Cambridge, Massachusetts. The remaining 3- and 4-year-olds were gathered from three nursery schools in Cambridge and Brookline, Massachusetts.

Design

Each child participated in six tasks: two weight tasks, two size tasks, and two density tasks (see Section 2). Because we wished to assess children's concepts uncontaminated by lexical bias, the three nonverbal tasks (size, weight, and density) were always run first, followed by the three verbal tasks. Within each age group equal numbers of children received the three nonverbal tasks in each of the six possible orders of presentation (e.g., size/weight/density; density/size/weight, etc.). The order in which the three verbal tasks were administered to each child was the same as for the three nonverbal tasks.

Stimuli

The same stimuli were used for the nonverbal size and weight tasks, namely, eight cubes varying in size and weight. The dimensions and weights of each cube are provided in Appendix 1. The four noncritical items included two cubes that were small and light (one made of aluminum and one of plexiglass) and two cubes that were both large and heavy (one made of wood and one of plexiglass). The four critical items included two cubes that were small but heavy (one made of steel and one made of brass) and two cubes that were large but light (one made of styrofoam and one of balsa wood). In addition there were four different cubes used for practice, two noncritical and two critical items. All of the practice cubes were painted blue; the eight test items were unpainted.

The box used in the nonverbal size task was a plastic cube open at the top: the small cubes fitted in it, the big cubes did not. The bridge used in the nonverbal weight task was made out of foam rubber. It collapsed under the weight of the heavy cubes but not under the light cubes.

The eight cubes were presented in one of four different orders, randomly assigned to an equal number of subjects. Each subject received the cubes in one order for the weight task and in a different order for the size task.

The stimuli for both the verbal size and verbal weight tasks were nineteen pairs of cylinders (these were also the stimuli for the verbal density task, see Section 2). The cylinders were made of wood, aluminum, or brass. The pairs fell into five different categories:

(1) $\underline{\mathrm{O}}\mathrm{o}$[1] (three pairs). The cylinders were made of the same material but were of different sizes. Size and weight covaried (noncritical items).
(2) $\underline{\mathbf{O}}\mathrm{o}$ (five pairs). The cylinders were made of different materials. The larger one was denser, so of course weighed more. Size and weight covaried (noncritical items).
(3) $\underline{\mathrm{O}}\mathbf{o}$ (two pairs). The cylinders were made of different materials. The larger one was less dense, but still weighed more. Size and weight covaried (noncritical items).
(4) $\underline{\mathbf{o}}\mathrm{o}$ (six pairs). The cylinders were the same size, but one was denser than the other, so heavier. Size and weight conflicted (critical items).
(5) $\underline{\mathbf{o}}\mathrm{O}$ (three pairs). The cylinders were different sizes and materials, such that the smaller, denser item weighed more than the larger, less dense item. Size and weight conflicted (critical items).

[1]The conventions for reading these diagrams are as follows: relative size is represented by relative size, the denser of the two items is in bold type, and the absolutely heavier of the two items is underlined.

In each verbal task, the child was presented with nine pairs: two $\underline{\mathrm{O}}\mathrm{o}$ pairs, one $\underline{\mathbf{O}}\mathrm{o}$ pair, two $\underline{\mathrm{O}}\mathbf{o}$ pairs, two $\underline{\mathbf{o}}\mathrm{o}$ pairs and two $\underline{\mathbf{o}}\mathrm{O}$ pairs. Three different series of nine pairs were prepared. Each child received all three series: one for the verbal size task, one for the verbal weight task, and one for the verbal density task (see Section 2).

Procedure

Children were individually tested at their schools by two experimenters. One conducted the experiment and the other noted the child's responses, spontaneous comments and relevant actions. Each child required several sessions to get through the six tasks. For the younger children, the battery was usually broken up into three 25-minute sessions on three different days. Older children usually required two 35-minute sessions.

Nonverbal size task

The task was preceded by a practice period during which the child was shown the box and the four painted practice items. A penny was placed at the bottom of the box. The experimenter asked the child to select a cube that would fit inside the box and touch the penny. (The penny was used to ensure that the child would interpret "fit inside" correctly.) Children were allowed to try out the cube they selected. The experimenter then asked the child to choose a cube that would not fit inside the box. Again, the child received feedback. The sequence was repeated for the remaining two practice cubes. This procedure was repeated if the child was incorrect about any prediction. In the second part of the practice, the four cubes were put into a bag, retrieved one at a time, and handed to the child. For each cube, the child was asked: "Will the block fit inside the box and touch the penny?" Younger children were directed to pick the cube from the bag themselves, to ensure that they would hold the cube and thus have information about its weight as well as its size before making their predictions. The procedure was repeated if any errors were made.

In the task proper, the eight unpainted test cubes were presented one at a time. The child was asked for a prediction in the form of a yes/no response and was allowed to check whether the cube did, or did not, fit inside the box.

Nonverbal weight task

The child was shown the bridge and a small plastic alligator who "lived under the bridge." The child was told: "People put blocks on top of the bridge, and some blocks make the bridge collapse and squash the alligator." The four practice cubes were brought out and the child was asked to choose one cube that would make the bridge collapse and squash the alligator. Next the child was asked to choose a cube that would not make the bridge collapse. This was repeated for the other two blocks and the entire sequence was run again if the child made any errors. The procedure for the practice and testing session of the nonverbal weight task was exactly parallel to the procedure for the nonverbal size task, except for the question asked. The child was asked about each block: "Will this block make the bridge collapse, and squash the alligator?" After each prediction, the child was allowed to put the block on the bridge to see what happened.

Verbal size task

Children were familiarized with the task by a series of questions: "Do you know that some things are larger than others and some things are the same size?" "Which is larger: a dog or an elephant?" "What is about the same size as an apple: an orange or a grape?"

Children were then given the nine pairs of cylinders, one pair at a time. While holding the two cylinders, they were asked: "Is one larger than the other, or are they the same size?" No feedback was given.

Verbal weight task

The warm-up questions were: "Do you know that some things are heavier than others and some things weigh the same?" "Who is heavier: you or your mom?" "What weighs about the same as an apple: an orange or a grape?"

The nine pairs of cylinders were then presented, one pair at a time. While holding the two cylinders, the child was asked: "Is one heavier than the other, or do they weigh the same?" No feedback was given.

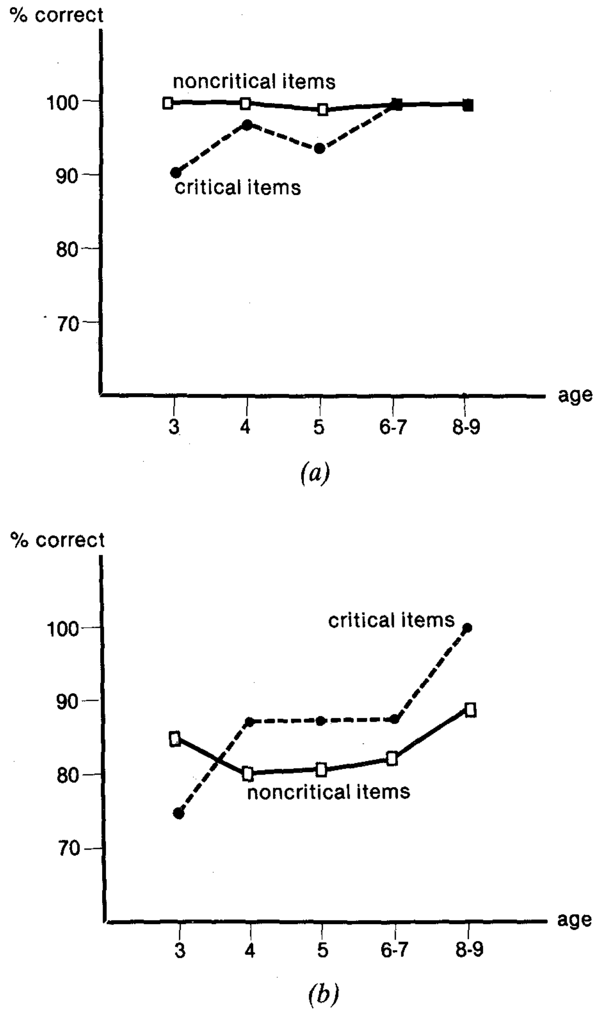
Results

Size tasks

Children's performance was excellent on both the critical and noncritical items (Figures 1a and 2a). Mistakes were extremely rare: 11 errors out of 624 responses on the nonverbal task (cubes in box) and, excluding a child who said "same" to all items, 13 errors out of 702 responses on the verbal task ("Is one larger ...?"). Every child age 6 and older had a perfect score on both tasks. So too did the majority (74%) of the 3–5-year-olds. Thus, most children gave direct evidence that their concept of size was fully differentiated from weight.

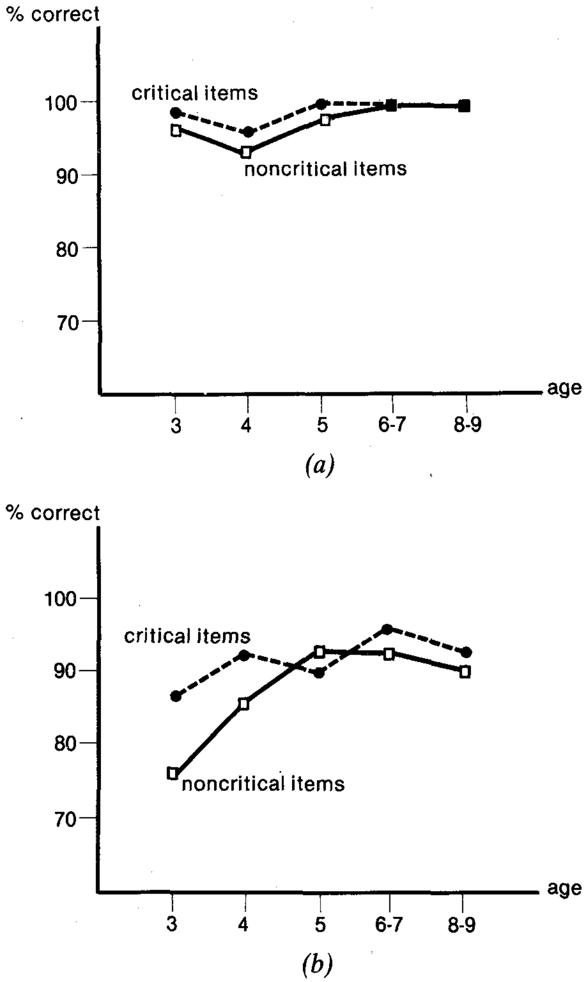
Seven children made errors on the nonverbal size task. These errors were

Figure 1. *Percent correct responses on the noncritical and critical items of the verbal size and weight tasks: (a) verbal size task; and (b) verbal weight task.*



evenly distributed among the critical and noncritical items. Nine children made mistakes on the verbal size task. Twelve of these 13 errors were on critical items. Thus, only on the verbal size task were the critical items significantly more difficult than the noncritical items. Most probably these errors are lexical in origin. Some children may have assigned a nondimensional meaning to "larger", by analogy to "more".

Figure 2. *Percent correct responses on the noncritical and critical items of the nonverbal size and weight tasks: (a) nonverbal size task; and (b) nonverbal weight task.*



Each child's responses were analyzed individually. A child's pattern was judged consistent with size/weight lack of differentiation ("size/weight" patterns) if the only errors were weight intrusions. Few children's responses met this criterion, and only 2 out of the 78 children had size/weight patterns on both size tasks (see Table 1).

Table 1. *Size and weight tasks: Number of children who show size/weight patterns*[a].

Age	Nonverbal size	Verbal size	Both size	Nonverbal weight	Verbal weight	Both weight	One size and one weight
3 ($N = 18$)	1	4	1	1	4	1	0
4 ($N = 18$)	1	2	1	1	1	0	0
5 ($N = 18$)	1	1	0	0	0	0	0
6–7 ($N = 12$)	0	0	0	0	0	0	0
8–9 ($N = 12$)	0	0	0	0	0	0	0
Total ($N = 78$)	3	7	2	2	5	1	0

[a] See text for definitions of size/weight patterns.

Weight tasks

Children made many more errors on the weight tasks than on the size tasks. Only 31 out of 78 children performed errorlessly on the nonverbal task (collapsing bridge) and only 25 of the 78 children performed errorlessly on the verbal task ("Is one heavier ...?").

Contrary to the *size/weight* nondifferentiation hypothesis, however, the error rate was lower on the critical than the noncritical items, both on the nonverbal tasks (10% vs. 14%) and the verbal tasks (14% vs. 16%) (see Figures 1b and 2b). The predominance of noncritical errors over critical errors is in fact larger than indicated by these numbers because over a quarter of the errors on the critical items were not size intrusions. For example, a size intrusion on the oO pairs would be to call the bigger one heavier; in fact, many children judged them to be the same weight. As a whole, then, children's difficulties were not attributable to *size/weight* lack of differentiation.

The errors on the nonverbal weight task cluster around two items: the small light aluminum cube (a noncritical item) and the small heavy brass cube (a critical item, see Table 2). To predict which cubes would make the bridge collapse correctly, the child must establish a cut-off point between heavy items and light items and hold onto it throughout the task. The small aluminum cube felt the heaviest of the four light cubes and the small brass cube felt the lightest of the four heavy cubes (see Table 2).[2] Those children

[2]Twelve adults provided magnitude estimations of the felt weights of the eight cubes. The cubes were presented one at a time. Subjects were asked to arbitrarily assign to the first cube a number corresponding to its felt weight, and then assign numbers to the succeeding cubes that respected the ratio of the felt weight of each one to that just judged. The items were presented to each subject in one of the four different orders used in the children's tasks. The data were normalized so that all the subjects' normalized scales had the same range, and the normalized felt weights were averaged across subjects.

Table 2. *Nonverbal tasks: Errors on individual items*

	Items							
	SmSteel	LgPlexi	LgWd	SmBr	SmAl	LgBal	SmPlexi	LgStyr
Weight (g)	440	760	790	260	80	127	60	42
Felt weight[a] (Adults)	101	77.9	65.1	48.5	10.7	9.2	7.7	3.3
Size task errors (Total)	3	0	0	1	3	0	1	3
Weight task errors (Total)	4	6	5	15	27	5	7	4

[a] See Note 2 for a description of how the felt weight measures were obtained.

who chose a different division into heavy and light items, that is between the large wood cube and the small brass cube or between the small aluminum cube and the large balsa wood cube, would produce a majority of errors on the small brass cube or small aluminum cube, respectively. Of the 38 children who made one of these two errors, only 4 made both, supporting the "different cut-off point" explanation. So too does the fact that these were the only two items on which adult subjects hesitated when predicting whether they would make the bridge collapse.

Table 3 shows the breakdown of errors for each pair type on the verbal weight task. The most striking feature of the error data is the prevalence, at every age, of saying "same" although the two cylinders were always of different weight. Overall, 89% of the errors were of this type. These errors are best interpreted as reflecting poor weight discrimination, rather than size/weight lack of differentiation, for several reasons. First, these errors were common on noncritical as well as critical items. However, only on the o̲o critical items is it even possible that "same" errors could reflect size intrusions. Second, within a particular pair type, the order of difficulty is predicted from adult psychophysical data[3] (see Appendix 2 for adult felt weight ratios). That is, within each pair type, the adults' ranking of the pairs according to how close in weight they felt was the same as the ranking according to the percentage of "same weight" judgments the children made.[4] Third, from age

[3]Ten adults served as subjects. They were instructed to close their eyes and were given the two cyclinders to hold. They were asked which cylinder was heavier and how many times heavier it was than the lighter cylinder. They were advised to base their judgments on felt weight only, ignoring any other information such as size or material. Their data were then normalized and averaged across subjects.

[4]The only exception was in the group of o̲o items. A brass/aluminum pair that had the least number of errors by the children was judged closest in weight by the adults.

4 to 9, the incidence of "same weight" errors was no higher on the critical items than on the noncritical items, and no child ever made a size intrusion on a oO critical item. Finally, although 3-year-olds were more likely to say "same weight" to the oo critical items than the noncritical items, these items were in fact closer in felt weight. Even 3-year-olds rarely judged the larger object to be heavier on the oO critical items. The group data, then, provides little evidence for size intrusions into weight judgments.

Each child's pattern of response was individually analyzed. For the nonverbal weight task, size/weight patterns were those in which errors were only on critical items, if those errors could not be explained in terms of a different cut/off between heavy and light items. For the verbal weight task, size/weight patterns were those in which the child picked the larger cylinder as heavier on at least one of the oO pairs, or if the child said "same" to at least one oo pair and made no other "same" errors. The number of children with size/weight patterns was small: only two had size/weight patterns on both weight tasks. Significantly these children were not the same children with size/weight patterns on the size tasks (see Table 1).

In sum, our data provide no evidence for size/weight lack of differentiation. In three of the four tasks, critical items were no harder than noncritical items. Further, children rarely had consistent size/weight patterns on individual tasks, and no child showed intrusions from one concept into judgments calling for the other on all four tasks. We conclude that 3- to 9-year-olds have distinct concepts of size and weight.

Table 3. *Verbal weight task: Percent errors by age and item type*

	Error type	Age				
Noncritical items	Incorrect reponse	3	4	5	6–7	8–9
Oo	Same weight	31%	14%	19%	21%	8%
	o Heavier	–	–	–	–	–
Oo	Same weight	14%	17%	22%	12%	17%
	o Heavier	–	6%	3%	12%	4%
Oo	Same weight	17%	11%	17%	17%	8%
	o Heavier	–	6%	–	–	–
Critical items						
oo	Same weight	44%	11%	19%	21%	–
	o Heavier	3%	–	–	–	–
oO	Same weight	14%	19%	6%	4%	–
	O Heavier	14%	–	–	–	–

2. The density tasks

Although the data in Section 1 indicate that children do not have an undifferentiated size/weight concept, it is easy to imagine what such a concept might be like. It is not so immediately obvious what a single weight/density concept might be like. The *density* we are considering is *heavy for size* and thus includes *weight*. How could the child have a density component without realizing that it defines a distinct concept from weight? The answer emerges when we note that heaviness is a concept with an implicit comparative structure. When one makes judgments of whether something is heavy or not, one must always make a comparison to some standard—as when one judges one object heavier than a particular other object, or when one judges an object heavy for a child but light for an adult. An object can be heavy for objects of that type (as in "heavy telephone") or heavy for objects of that size. The child might have a concept of weight which includes the component *heavy for size* along with other components that determine relevant standards for comparison. Certain contexts might make one component more salient than the others. But until children realize that the component *heavy for size* defines a distinct physical magnitude relevant for making comparisons of the densities of material, they cannot be said to have distinguished the concepts of weight and density.

Piaget and Inhelder (1974) did not credit children with having distinguished *weight* from *density* until they could explain the density differences of materials in terms of schemas of compression and decompression within a particulate theory of matter. This criterion seems too strong. In the history of science, the concept of density of materials was present early, long before atomistic conceptions of matter were accepted. We credit children with the concept of density if they realize that they have to use a different sense of weight in comparing the heaviness of objects and in comparing the heaviness of materials. If children have a single weight/density concept, they should make density intrusions on weight judgments and weight intrusions on density judgments. Finally, if *density* is absent from their conceptual system, they should always fail to relativize to size when making either weight or density judgments.

Four tasks—two verbal and two nonverbal—probed for *weight/density* lack of differentiation. The verbal weight task was the same task as described in Section 1 with the data reanalyzed relative to a new definition of critical items (those relevant to *density/weight* rather than *size/weight* differentiation). In the verbal density task, we presented children with the same pairs of cylinders and asked "Is one made of a heavier kind of stuff or are they made of the same kind of stuff?" We would not expect 3- and 4-year-old children to know the meaning of "dense" or be able to understand any locution such as "heavy for its size" (see Quintero, 1980). Extensive pretraining was given as to what was meant by "heavier kind of stuff." In both the verbal weight and verbal density tasks, critical items are those in which the heavier object in a pair is not made of a denser material (i.e., density and weight conflict).

In the first nonverbal density task, we gave children pretraining designed to teach them that steel objects are heavier for their size than aluminum objects. We then asked children to sort new objects one by one into steel and aluminum families. The second nonverbal density task[5] required them to decide which of two objects was steel and which was aluminum. On both tasks, critical items were those where a steel object is lighter than an aluminum object.

Method

Stimuli

The test stimuli for the verbal weight and verbal density tasks were the same cylinders that had been used in the verbal size and weight tasks (see Appendix 1). There were no warm-up stimuli for the verbal weight task. The stimuli for the warm-up session in the verbal density task were three blocks which were painted blue (one large styrofoam block, one small styrofoam block, and one small aluminum block) and three small unpainted cylinders of equal size (one brass, one wood, and one aluminum). Although painted blue, it was easy to see that the styrofoam blocks were made of something different from the aluminum block.

The stimuli for the nonverbal density task consisted of 12 practice items and 10 test items. The practice items included 6 pairs of steel and aluminum pieces, each pair being the same size. Two of the practice pairs were uncovered; the other 4 pairs were covered with yellow contact paper. The 10 test stimuli (5 made of aluminum, 5 made of steel) were also covered with yellow contact paper. In both the practice and test stimuli, some of the largest aluminum pieces were heavier than the smallest steel pieces (critical items: see Appendix 1 for details). Stimuli for the forced choice task were the 2 largest aluminum test blocks and the 2 smallest steel blocks. In each pair, the aluminum item was (and felt) absolutely heavier than the steel item. That is, the 2 forced choice pairs were both critical items for diagnosing weight/density lack of differentiation.

[5]The data from the nonverbal weight task could not be used to document weight/density lack of differentiation because there were no items in which density and weight clearly conflicted.

Procedure

Verbal weight task

The procedure for the verbal weight task was described in the previous section.

Verbal density task

The verbal density task began with a warm-up. Three blue blocks (one large styrofoam, one small styrofoam, and one small aluminum) were placed in front of the child. The child was told that "Some things are made of different kinds of stuff; some things are made out of heavier kinds of stuff than others." The tester then asked: "Are any of these made of the same kind of stuff? Which one is made of a heavier kind of stuff?" Incorrect responses were corrected and discussed. Next children were given brass, wood, and aluminum cylinders which were all the same size. They were told the name of the material each was made of and were allowed to pick them up. The tester then paired together the brass and aluminum piece and asked: "Which is made of a heavier kind of stuff, or are they both made of the same kind of stuff?" Then the same procedure was followed with the aluminum and wood pieces. Any incorrect responses were corrected and discussed.

In the test phase, the child was handed a pair of items, one item in each hand. The tester then asked: "Is one of these made of a heavier kind of stuff, or are they made of the same kind of stuff?" There were two types of critical items ($\underline{\mathrm{O}}\mathrm{o}$ and $\underline{\mathrm{O}}\mathbf{o}$) in which the heavier item was not made of the denser material, and three types of noncritical items ($\underline{\mathbf{O}}\mathrm{o}$, $\underline{\mathbf{o}}\mathrm{o}$ and $\underline{\mathbf{o}}\mathrm{O}$) in which weight and density covaried.

Nonverbal density task

Preliminary warm-up and practice. The extensive practice session began with the experimenter handing the child two large, uncovered cylinders and telling the child which one was made of steel and which one was made of aluminum. Next, two small steel and aluminum cylinders were handed to the child, who was asked which was steel and which was aluminum. The child was then asked to put the two steel cylinders together in a steel family and the two aluminum cylinders together in an aluminum family. Naming and sorting mistakes were corrected. The experimenter then presented the child with four same size pairs of steel and aluminum that were covered with yellow contact paper. For each pair, the experimenter said: "Which is steel and which is aluminum? Put the steel one in the steel family and the aluminum one in the aluminum family." If any errors were made, the previously sorted items were again sorted. Finally, all of the covered blocks were then mixed up and given to the child one at a time to sort into families. Any pieces that the child sorted incorrectly were presented again, until correctly placed.

Sorting into steel and aluminum families. The child was asked to sort the ten covered test blocks into two groups: a steel family and an aluminum family. The families were marked for the child by placing an uncovered piece of each material in place on the table. If the child said "steel" but placed the item in the aluminum family (or vice versa), the discrepancy was pointed out and the child could change his or her placement. The final placement of the piece determined the scoring. Occasionally, children compared the item handed to them with previously sorted items before deciding where to place it. This was allowed because it demonstrated awareness of the problem posed by the varied sizes of the pieces. These actions, as well as any spontaneous justifications, were recorded.

Forced choice task. After all ten pieces had been sorted into steel and aluminum families, the child was given the two forced choice problems. The tester handed the child one of the pairs and said: "One of these is steel and one is aluminum. Which is steel?" After each choice, the children were asked how they could tell which piece was steel and their comments were recorded.

Results

The nonverbal density tasks

Sorting into steel and aluminum families. At every age, the error rate for the critical items far exceeded that for the non-critical items (Figure 3). Overall, 85% of the errors were missorts of the critical items.

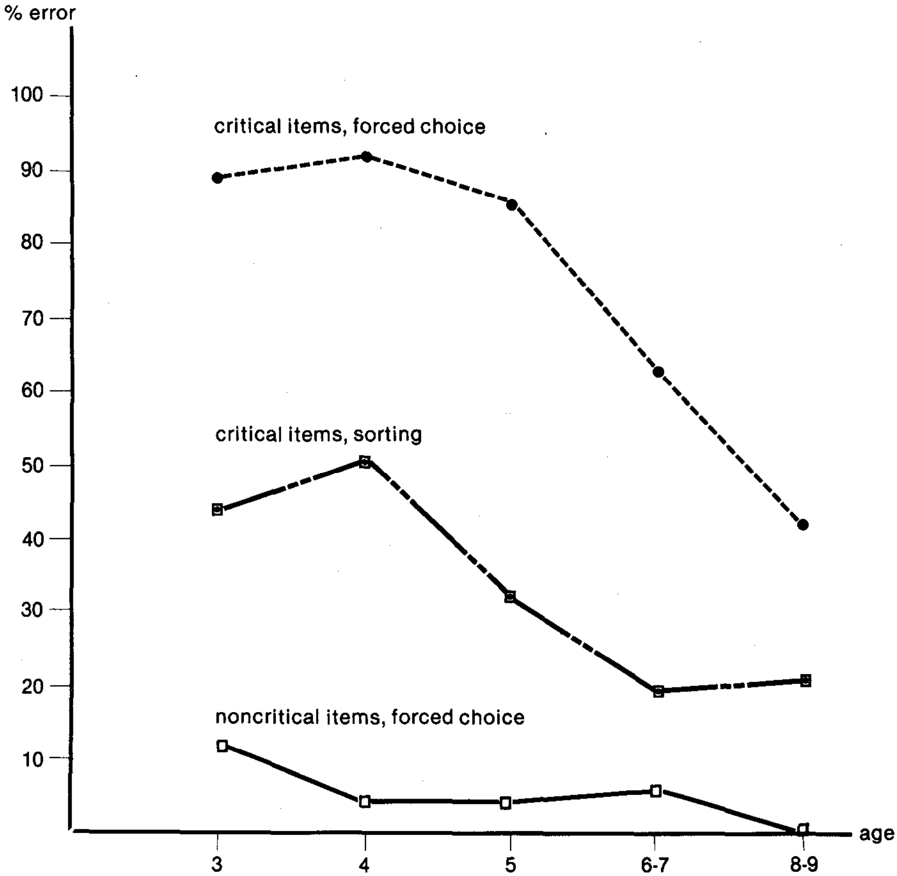
Analysis of each child's pattern of judgments posed some difficulty. Perfect sorting is a clear density pattern, while sorting the four critical items according to felt weight is a clear weight pattern. However, the most common patterns involved at least one critical item correctly sorted and at least one critical item incorrectly sorted. Depending upon which items were missorted, these patterns might either be entirely consistent with weight judgments, or they might reflect at least one judgment where size had been adjusted for. The reason for this ambiguity can be seen when subjects' responses to the stimuli are arranged according to the stimuli's felt weights (see Table 4). The ordering of the felt weights of the ten stimuli was obtained from magnitude estimations provided by adult subjects.[6] Children's patterns were judged to be

[6]The magnitude estimations of adult subjects were normalized to 1 for 1 Al. The numbers for other items reflect the ratio of felt weight estimations between them and 1 Al, e.g., 3 Al was judged 6.5 times as heavy as 1 Al. Similarly, 4 Al was judged 12.5 times as heavy as 1 Al, making 4 Al almost twice as heavy as 3 Al.

weight patterns whenever there was a break between heavy and light items that corresponded to their classification into aluminum and steel families. Patterns 1 and 2 in Table 4 are both weight patterns. In Pattern 1 the break between the aluminum (Al) and steel (St) families comes between 2 St[7] and 4 Al, resulting in four critical items being missorted. In Pattern 2 the break comes between 5 Al and 3 St, resulting in only two critical items being missor-

[7]As explained in Table 4, 1 St is the smallest steel block, 2 St the next smallest, and so on, up to 5 St. Similarly, 1 Al is the smallest aluminum block, and 5 Al the largest one.

Figure 3. *Percent errors on the nonverbal density tasks: noncritical versus critical items on the sorting into steel and aluminum families task and critical items on the forced choice task.*



ted. Nonetheless, the child's missorting perfectly respects felt weight. Pattern 3 is the perfect density pattern (no errors). Children were also credited with a density pattern if they made one error that appeared to be a slip-up, because it did not involve a critical item (e.g., Pattern 4). Finally, children's patterns were judged a mixture of density and weight judgments (density/weight) if they made at least one error on a critical item and were correct on at least one critical item, and if their pattern of judgments was not consistent with a single felt-weight cut-off. Patterns 5, 6, and 7 are three examples. In Pattern 5, two critical items were assigned properly (1 St and 4 Al) and two were not (2 St and 5 Al). The two that were misclassified seem to reflect weight intrusions; there is a break between light and heavy items between 4 Al and 5 Al. However, assigning 1 St to the steel family is not consistent with a pure weight pattern, as it is far lighter than three items placed in the aluminum family. For this item, the child must have adjusted for the smallness of the

Table 4. *Nonverbal density task: Patterns of responding in the sorting task*

	Items[a]									
	1 Al	2 Al	1 St	3 Al	2 St	4 Al	5 Al	3 St	4 St	5 St
Weights (g)	21	54	72	164	148	263	300	441	715	850
Felt weight estimations (adults)	1	2.5	4	6.5	8.5	12.5	17	21	49	52.5
Types of patterns										
Pattern 1 (Weight)	Al	Al	Al[b]	Al	Al	St	St	St	St	St
Pattern 2 (Weight)	Al	Al	Al	Al	Al	Al	Al	St	St	St
Pattern 3 (Density)	Al	Al	St	Al	St	Al	Al	St	St	St
Pattern 4 (Density)	Al	Al	St	Al	St	Al	Al	Al	St	St
Pattern 5 (Density/Wt)	Al	Al	St	Al	Al	Al	St	St	St	St
Pattern 6 (Density/Wt)	Al	Al	Al	Al	St	Al	Al	St	St	St
Pattern 7 (Density/Wt)	Al	Al	St	Al	St	Al	St	St	St	St

[a] 1 Al is the smallest aluminum cylinder; 5 Al, the largest.
1 St is the smallest steel cylinder; 5 St, the largest.
[b] Circled responses are errors.

block. Similarly, in Patterns 6 and 7 only one critical item is missed, and it is consistent with a weight intrusion. However, the other critical items are correct and required some size adjustments.

The responses of five of the youngest children could not be classified into any of these pattern types. But most 3- and 4-year-olds had systematic patterns, showing they could at least sort according to weight (Table 5). The modal pattern for the 5–7-year-olds was the mixed density/weight pattern. Not until ages 8–9 did density patterns become modal.

Analyses of individual patterns, thus, confirm the group results shown in Figure 3. At all ages weight intrusions were the predominant kind of errors, and only at the oldest ages could some children reliably correct for size in sorting the blocks into the steel and aluminum families. The pattern analyses also add a new dimension to the results. They show that before age 5 few children make any judgments based on *heavy for size*, despite the extensive pretraining. Thus, at this early age children's weight concept may have no component of density. Between ages 5 and 7, however, children frequently make some judgments based on *heavy for size* and some on *heavy*. These children are good candidates for having an undifferentiated weight/density concept.

The forced choice task. Both comparisons in the forced choice task involved critical items, in the sense that the object made of the heavier kind of stuff (steel) weighed less than the object made of the lighter kind of stuff (aluminum). As can be seen from Figure 3, children at every age made many

Table 5. *Nonverbal density tasks: Pattern analyses for the sorting and forced choice tasks*

Age	Sorting task				Forced choice task		
	Density	Density/ weight	Weight	Other	Density	Density/ weight	Weight
3 ($N = 18$)	11%	6%	61%	22%	6%	11%	83%
4 ($N = 18$)	0%	33%	61%	6%	6%	6%	89%
5 ($N = 18$)	22%	50%	28%	0%	0%	33%	67%
6–7 ($N = 12$)	42%	58%	0%	0%	33%	8%	58%
8–9 ($N = 12$)	50%	25%	25%	0%	33%	50%	17%

errors on these problems. The 3- and 4-year-olds virtually always missed them, and the oldest children missed almost half of them.

Since there were only two forced choice comparisons, the possibilities for within child pattern analyses were limited: the child could base both comparisons on weight, calling the steel blocks aluminum and the aluminum blocks steel (weight pattern), the child could base both comparisons on density (density pattern), or the child could base one comparison on density and one on weight (density/weight pattern). Table 5 shows there was a regular decrease with age in weight patterns. Also, density/weight patterns were quite common, indicating that some children managed to adjust for size in one, but not both, of their judgments.

In broad outline, the developmental stories that emerge from the families task and the forced choice task are the same. Both tasks show that the youngest children relied mainly on weight, both tasks provide some density/weight patterns, and in both tasks pure density patterns are observed primarily in the two oldest age groups. However, there was a major difference in the data between the two tasks. Figure 3 shows that at every age weight intrusions were almost twice as frequent on the forced choice task. Given that the critical items on the two tasks were exactly the same blocks (1 and 2 St, 4 and 5 Al), why was the forced choice task so much harder than the task of sorting the items, one at a time, into steel and aluminum families?

Two factors might have conspired to make the forced choice items more difficult. First, the explicit comparison of two blocks, one placed in each hand, might have made absolute weight differences particularly salient. Second, the reasoning underlying a correct judgment in the forced choice task was complex. Children had to note that while the larger item was heavier than the smaller one, it was not heavier *enough*, given the vast difference in size between them, for it to be made of the heavier kind of stuff. While children could reason along these lines in the families task, another strategy was also available to them. Children might compare the weight of a block to one of the immediately preceding blocks of roughly the same size. Often children encountered one of the small aluminum blocks before encountering the first of the small steel blocks. Given that the weight of the steel block was much greater than that of another block close in size, they might have concluded that it must be steel. Similarly, children often encountered one of the large steel blocks before one of the large aluminum blocks. Relative to those large steel blocks, the large aluminum blocks were light. While this strategy involved taking size into account in judging whether an object was steel or aluminum, a judicious choice simplified the inference. The child simply needed to reason: the sizes are about equal; one is much heavier; therefore, it must be steel. In sum, we are suggesting that the information processing demands of the two tasks differed: the computations involved in the forced choice task were more difficult than those in the families task.

There is a direct way to test whether children with density/weight patterns were using this strategy in the families task. If they were, they should have done better on the 1 and 2 St items when these items closely followed 1 or 2 Al items than when many other items intervened. This is because these items are about the same size, and children can only use the above strategy when items similar in size (but different in materials) closely follow one another. Similarly, they should have done better on the 4 and 5 Al items if those items had closely followed 3, 4, or 5 St than if many other items had intervened. Table 6 shows that this is the case. When critical items were correctly sorted, there was on average a relevant contrast 1.8 items earlier. In contrast, when critical items were incorrectly sorted, relevant contrasts appeared on average 3.1 items earlier (T = 36, N = 25, $p < .005$, Wilcoxon signed ranks test, 1-tailed). This relation holds at every age, for the youngest two groups (T = 3, N = 7, $p < .05$, Wilcoxon signed ranks test, 1-tailed) as well as the oldest three (T = 21, N = 18, $p < .005$, Wilcoxon signed ranks test, 1-tailed). One would not expect this relation to hold for the children with pure weight patterns, since their patterns give no evidence of their ever taking size into account. Indeed, it does not (see Table 6). The average number of items intervening between a correctly sorted critical item and its closest relevant contrast is 2.4; between an incorrectly sorted critical item and its closest relevant contrast, 2.5.

This analysis supports our hypothesis about why the families task was easier than the forced choice task. The families task afforded relevant contrast items making it easier for the children to relativize weight to size when judging kind of stuff. The analysis also confirms that those children whose patterns were classified as density/weight were indeed different from those whose patterns were classified as pure weight. Only the former took advan-

Table 6. *Nonverbal sorting task: Average distance from informative contrast*

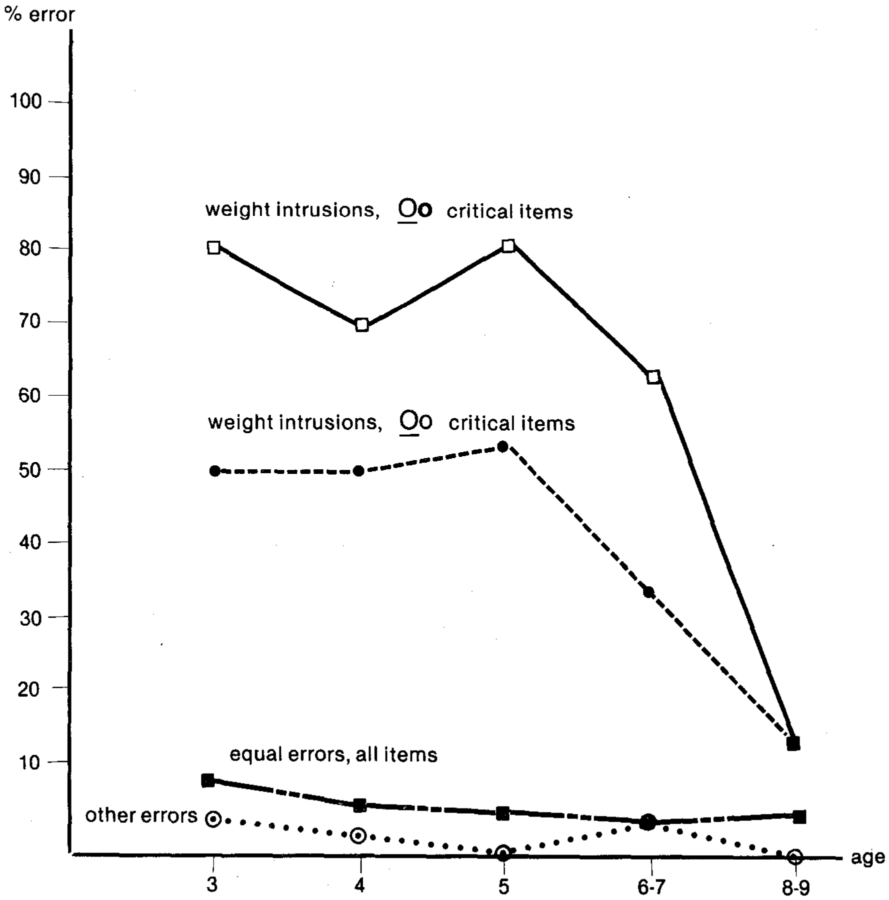
	Density/weight pattern		Pure weight pattern	
Age	Correctly sorted items	Incorrectly sorted items	Correctly sorted items	Incorrectly sorted items
3–4	1.6	2.9	2.6	2.5
5–9	1.9	3.2	2.0	2.6
3–9	1.8	3.1	2.4	2.5

tage of the relevant contrast; that is, only they were sometimes judging heaviness for size.

Verbal density task

Virtually all the errors, at every age, on the verbal density task were weight intrusion errors (see Figure 4). There were two types of errors. Most commonly, children judged that the heavier cylinder in a critical item pair (Oo and **O**o) were made of the heavier kind of stuff. This is a clear weight

Figure 4. *Percent errors on the verbal density task: weight intrusions on the critical items, equal errors (possible weight intrusions) on all the items, and other errors on the noncritical items.*



intrusion. Much less commonly, objects made of visibly different kinds of stuff were judged to be made of the same kind of stuff. Errors of this sort were made for both noncritical and critical items. Given that on the verbal weight task the dominant error was to say that the items weighed the same (because of poor felt weight discrimination—see Section 1), these "same kind of stuff errors" are probably weight intrusions as well. Thus, in broad outline, the group data from the verbal density task tell the same story as those from the nonverbal density tasks.

An analysis of individual patterns revealed that there were four main patterns: (1) density patterns (perfect scores on both critical and noncritical items); (2) density/weight patterns (one judgment on a **Oo** item based on absolute weight, and one based on density); (3) weight and kind of stuff patterns (both **Oo** critical items were judged on the basis of weight, while the Oo critical items were correctly judged as being made of the same kind of stuff); and (4) weight patterns (all items, including the four critical items, judged on the basis of absolute weight; "same kind of stuff" errors made on items of visibly different materials were judged as weight intrusions as well). All of the older children's patterns could be classified with this scheme; a third of the 3-year-olds's patterns could not be classified.

Table 7 shows the distribution of patterns with age. Two of the patterns reveal no knowledge of density in the sense of heavy for size: the pure weight pattern and the kind of stuff/weight pattern. Children showing the latter pattern needed only to note from the appearance of the materials that the Oo critical items were made of the same kind of stuff. They never needed to consider that both objects were equally heavy for their size. Almost all the systematic patterns of the 3–5-year-olds were of these two types. Not until ages 6–7 did heavy for size become a major factor in the child's responses.

Table 7. *Verbal density task: Pattern analysis*

Patterns	Age				
	3 (N= 18)	4 (N = 18)	5 (N = 18)	6–7 (N = 12)	8–9 (N = 12)
Other	33%	0%	0%	0%	0%
Weight	28%	56%	56%	42%	17%
Kind of stuff/ weight	22%	28%	33%	17%	8%
Density/weight	0%	11%	11%	8%	0%
Density	17%	6%	0%	33%	75%

Finally, three-quarters of the 8–9-year-olds provided perfect density patterns.

The number of density patterns on the verbal density task may be compared to the number of density patterns on the forced choice task. In order to have a density pattern on each task, the child had to reason that the bigger, heavier item was not heavier *enough* to be made of the heavier kind of stuff. However, the objects used in the verbal density task were uncovered and children had been given information in the warm-ups that a piece of brass is heavier than a same size piece of aluminum and a piece of aluminum is heavier than a same size piece of wood. Thus, unlike in the forced choice task, in the verbal density task children can use knowledge of the visual differences in the materials to aid them in making judgments. If children understand density, but have difficulty correctly relativizing weight judgments to size, one would expect better performance on the verbal density task than the forced choice task. The number of density patterns on the two tasks were comparable at every age, except among the oldest children. Among these children, 75% achieved density patterns on the verbal density task, compared to 33% on the steel and aluminum forced choice task ($p < .05$, Fisher exact test, 1-tailed). Apparently, then, only the oldest children were aided by having the materials clearly visible.

There was one further respect in which patterns of performance on the verbal density tasks differed from patterns on the nonverbal density tasks. Density/weight patterns were common on both nonverbal density tasks but rare on the verbal density task. Why should this be? In the nonverbal density tasks, children had been given extensive pretraining making the *heavy for size* component of their weight concept more salient (i.e., they had to sort critical and noncritical items). In contrast, no critical items (e.g., items in which the lighter object was denser) were presented in the pretraining of the verbal density task. Children were presented with large and small cubes made of styrofoam and a small cube made of aluminum. They were asked to judge which objects were made of the same kind of stuff and which were made of a heavier kind of stuff. They were then introduced to the three materials used in the verbal density task by seeing and lifting three cylinders of equal size: one made of wood, one of aluminum, and one of brass. They were asked for a given pair "Is one of these objects made of a heavier kind of stuff, or are they made of the same kind of stuff?" Children who already had a concept of density could infer from this experience that brass objects are heavier for their size than aluminum objects and aluminum objects are heavier for their size than wooden ones. However, children who did not have a distinct concept of density might only note that the brass object was heavier than the aluminum object or infer that brass objects are heavier than aluminum ones. For these children, then, the preliminaries would not have activated their notion of *heavy for size*, although they could have made their notion of kind of stuff more salient. Thus, the typical patterns on the verbal density task among children with an undifferentiated weight/density concept would be weight and kind of stuff/weight patterns instead of density/weight patterns. This is what we observed.

Pattern analysis for density tasks

An overall pattern analysis assessed the consistency of children's responding across the three density tasks (Table 8). Half of the 3- and 4-year-olds judged only on the basis of absolute weight, kind of stuff and weight, or were unclassifiable on at least one task. These children gave no evidence that they had *heavy for size* at all, even undifferentiated from heaviness. Most of the 5-year-olds and 6–7-year-olds had overall patterns consistent with *weight/density* lack of differentiation. There were two such overall patterns. In the most common weight/density lack of differentiation pattern (type 1: 35 of the 46 cases), children relativized weight to size on some but not all of the nonverbal density judgments, but then produced weight patterns, or weight/kind of stuff patterns, on the verbal density task (type 1 patterns). This pattern indicates that these children had a *heavy for size* component to their weight concept, but they used it primarily in contexts where it was made salient by pretraining or by the structure of the task (e.g., sorting objects one at a time). The fact that for these children the verbal density pretraining did not make this component salient indicates that they did not spontaneously realize that this is the sense of weight called for in making inferences about the densities of materials. The other children with an overall weight/density pattern (type 2: 11 of the 46 cases) had a perfect density pattern (or in a few cases, a density/weight pattern) on the verbal density task, but then a pure weight pattern on the forced choice aluminum and steel task, and a weight or density/weight pattern on the families task. These children were clearly more sophisticated when the materials could be visibly identified than when they had to make inferences solely from the objects' sizes and weights. They might have been following a strategy in the verbal density task based solely on the identification of the materials. Since they had to understand the preliminary warm-ups in the verbal density task to adopt this strategy, we must credit them with some understanding of heavy for size. However, they seem to judge primarily on the basis of felt weight when the materials are not visible. The extensiveness of the weight intrusions on the nonverbal density tasks make them hard to dismiss as occasional information processing slip-ups. Instead, children seem not to realize that they must relativize weight to size in this context. Finally, most of the older children demonstrated a notion of density differentiated from weight. Some of these children were perfect on all three density

Table 8. *Density tasks: Overall pattern analysis*

Pattern	Age				
	3 (N= 18)	4 (N = 18)	5 (N = 18)	6–7 (N = 12)	8–9 (N = 12)
Pure weight or other	50%	50%	17%	–	–
Undifferentiated weight/density					
Type 1	33%	33%	77%	50%	25%
Type 2	17%	17%	6%	25%	9%
Density					
Density consistent	–	–	–	17%	33%
Perfect density	–	–	–	8%	33%

tasks. Others were perfect on the verbal density task and only occasionally had weight intrusions on the nonverbal density tasks. Their difficulty may simply have been in correctly relativizing weights to size on the basis of felt weight cues. The fact that they correctly did so on at least one of the forced choice aluminum and steel items suggests that they had some understanding of the basic structure of the forced choice and verbal density tasks.

Verbal weight task

As described in Section 1, children made many errors in the verbal weight task, but the main source of these errors was poor weight discrimination. The major error was saying that the two cylinders weighed the same amount when in fact they differed, and could easily be seen to differ by adults. Poor sensory discrimination could not be the only cause of error, however, because adult psychophysical ratings were not predictive of children's error rates across all pair types. The overall rank correlation coefficient between adults' psychophysical ratings and children's error rates was nonsignificant (r = .4, df = 14). Therefore, the specific characteristics of each pair type influenced the children's judgments. We showed in Section 1 that size/weight lack of differentiation did not contribute to the responses. Two aspects of the data summarized in Table 3 suggest that density intrusions, in contrast, did play a role in the 4–9-year-old children's errors. First, one kind of item in which weight and density conflicted, the **O**o pairs, were especially difficult (23% errors), although they were not judged particularly close in felt weight by our adults. Second, these were the only items to lead to non-"same" errors: that is, the child sometimes picked the small denser cylinder as the heavier. In contrast, the Oo items (in which density and weight covaried) led only to "same" errors, and yielded a much lower error rate among 5–9-year-old children.

Each child's individual pattern of response was analyzed. A pattern was considered a weight/density pattern if the child picked the denser object as heavier on at least one of the Oo pairs, or if the child said "same" to at least one of the Oo pairs, and made no other "same" errors. Eleven children met these criteria. Of these 11 children, 8 had also been classified as using an undifferentiated weight/density concept in the density tasks.

If some children have an undifferentiated weight/density concept, why should there be fewer density intrusions into weight judgments than weight intrusions into density judgments? One possibility is that the verbal weight task was accompanied by no pretraining that made the *heavy for size* component of their weight concept salient. Also, this task required children to compare two objects placed in their hands—a context which may make differences in felt weight more salient than does sorting objects one at a time. Making judgments solely on the basis of absolute weight can be consistent with an undifferentiated weight/density concept, so long as the context has made absolute weight differences more salient than weight differences relativized to size. Of the 8 children with a weight/density pattern in the verbal weight task who had been classified as using an undifferentiated weight/density concept in the density tasks, 7 had had a density task earlier in the same session, in which they had made some heavy for size judgments. Thus, these children made their verbal weight judgments in a context that made the *heavy for size* component of their undifferentiated weight/density concept relatively salient.

In sum, the data concerning *weight/density* lack of differentiation contrasts markedly from those concerning *size/weight* lack of differentiation. In Section 2, we see massive intrusions of weight into density judgments, and some intrusions of density into weight judgments. Some of these errors resulted from apparent absence of the concept of density, but many children provided patterns across several tasks consistent with an undifferentiated weight/density concept.

3. The relation between size and weight

The children in Piaget's and Inhelder's studies equalized sizes when asked to make a clay ball that weighed the same as a given wax ball. The data from Section 1 suggest that this size intrusion into a weight judgment could not be due to *weight* being absent from the child's conceptual system, or from *size/weight* lack of differentiation. The first goal of the study to be reported in

Section 3 is to check that the children in our sample would make this same error. If so, we will conclude that the error results from an incorrect conceptualization of the relation between size and weight.

A second goal is to explore further how children conceptualize the relation between size and weight. A mature conceptualization of this relation requires an understanding of density, since the weight of an object is determined by both its size and the density of the material it is made of. In Section 2, we found that most 3–7-year-old children did not have a concept of density distinct from weight. How could these children conceptualize the relation between size and weight? They might have no generalizations relating size and weight. Alternatively, they might believe that an object's size is a rough predictor of an object's weight. Such a belief could arise as a crude empirical generalization from everyday experience. These children might have another crude empirical generalization as well: the kind of material an object is made of also is a rough predictor of an object's weight. Finally, some children might expect that size and weight perfectly covary and have no knowledge of the role of materials. This belief may even be theory driven. If the child thinks weight is proportional to amount of stuff, and does not know that different kinds of stuff have different densities, his beliefs dictate that size and weight be perfectly correlated.

Method

Subjects

The same children tested in the previous sections were the subjects in this study.

Procedure

We devised a variant of Piaget's and Inhelder's wax and clay ball task. Our problems involved two different sets of materials—steel versus aluminum and wax versus clay. The steel/aluminum problems were always presented first, because the children's extensive experience with these materials in the nonverbal density task might help them on these problems. The steel/aluminum problems were presented in a new session after the six tasks discussed in Sections 1 and 2. Older children completed the whole procedure, including the wax/clay problems, in one session; younger children required two 30 minute sessions.

In the preliminary session, the experimenter began by placing a large aluminum object and a small steel object of equal weight on the table in front of the child (O=o). The child was asked: "Is one of these pieces heavier than the other or do they weigh the same?" It was noted whether the child spontaneously chose to feel the object before answering; judging the bigger one heavier would be consistent with the belief that size perfectly predicts weight. The objects were placed in the hands of those who did not spontaneously pick them up, and the question was asked again. Because of the size/weight illusion, many of the children judged the small piece of steel to be the heavier of the two. The two objects were then placed on a balance to show that they weighed the same and the children were asked a final time whether one object was heavier than the other or whether they weighed the same. At this point, children were also asked two questions to assess their thinking about density differences: "Are these two objects made of the same kind of stuff or are they made of different kinds of stuff? Is one of these objects made of a heavier kind of stuff?" The experimenter then brought out two new pieces of steel and aluminum which were both the same size (oo). Children were first asked to predict whether these two objects would weigh the same or whether one would be heavier. This tests whether they understood the implications of the previous demonstration. Finally, they were allowed to lift the objects and put them on the balance to see that the steel object was in fact heavier. The objects were left on the balance throughout the Piagetian tasks as a reminder that the steel piece was heavier than the aluminum piece.

These two preliminary problems were followed by three forced choice problems involving steel/aluminum comparisons. Children were given a target object of one material. Their task was to pick one of three objects made of the other material which was equal in weight to the target. On each problem, one of the choices was the same size as the target. On one of the three problems, the same size piece was the largest of the three choices, to distinguish a same size pattern of responding from a pattern consistent with always picking the middle-size object. Two of the three problems had a steel object as target, while the other had an aluminum object as target. This ensured that children could not be systematically correct simply by picking the largest object.

The wax and clay problems were the same, except for two modifications. One question was added at the end of the preliminary problems. After they saw that a clay ball did not balance with a wax ball the same size, children were asked: "If these are the same size, why don't they weigh the same?" This question assessed whether children would articulate that the kind of material an object was made of affected its weight. Second, after children made their first wax/clay forced choice judgment, they were challenged to see whether they were able to resist a counter-suggestion. Children who chose a different size piece were told that another child had thought the wax and clay balls had to be the same size in order to weigh the same. Children who had picked the same size alternative were told that another child thought the wax ball had to be larger in order to weigh the same. Following these challenges, children were given the last two forced choice comparisons to determine if the challenge had affected their pattern of judgments.

Results

The preliminaries

Thirty-eight percent of the children were willing to make initial predictions on the steel and aluminum problems. Overwhelmingly, these children predicted on the basis of material rather than size (see Table 9). When confronted with less familiar materials (i.e. wax and clay), the majority (78%) wanted to lift the objects before venturing an answer. Only a few (13%) made initial predictions on the basis of size. Thus, even in the absence of knowledge of specific materials, children gave little evidence of believing size was a good predictor of weight.

The preliminary questions also confirmed young children's lack of under-

Table 9. *Steel/aluminum and wax/clay preliminaries: Percent of children lifting objects or making predictions of a particular type*

	Age				
Type of problem and answer	3 (N= 18)	4 (N = 18)	5 (N = 18)	6–7 (N = 12)	8–9 (N = 12)
Steel/aluminum preliminary ($O_=o$)					
Lift first	50%	67%	77%	58%	50%
Predict same weight	11%	–	6%	–	8%
Predict steel heavier	22%	33%	17%	25%	42%
Predict aluminum heavier	17%	–	–	17%	–
Wax/clay preliminary ($O_=o$)					
Lift first	67%	83%	88%	67%	83%
Predict same weight	17%	–	–	–	8%
Predict clay heavier	6%	6%	6%	–	–
Predict wax heavier	11%	11%	6%	33%	8%

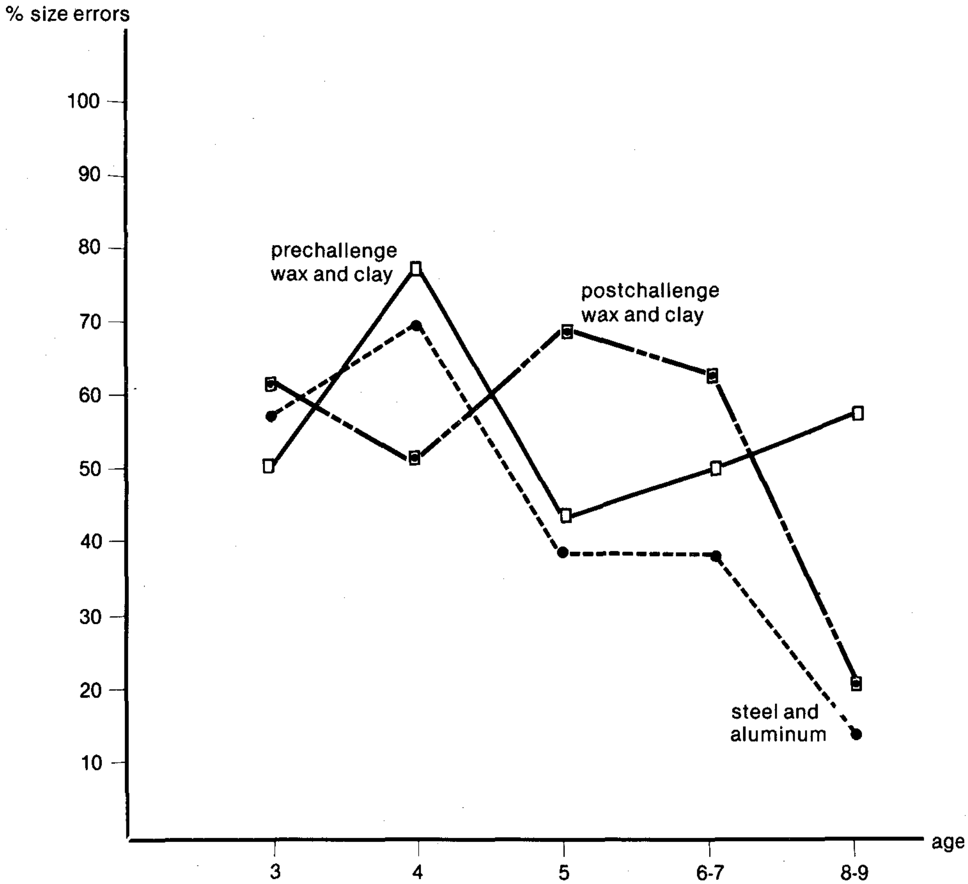
standing of density. After children were shown that a large aluminum piece weighed the same as a smaller steel piece, they were asked whether the objects were made of the same kind of stuff, and if not, whether one of the objects was made of a heavier kind of stuff. Children were then shown two new pieces of steel and aluminum, both the same size, and asked whether they would weigh the same. Only a few of the 3- and 4-year-olds concluded that steel is a heavier kind of stuff than aluminum and then predicted that the steel object (in the $\underline{\mathbf{o}}\mathbf{o}$ pair) would be heavier. Frequently children were inconsistent (e.g., initially concluding that steel is a heavier kind of stuff, but then predicting that the two $\underline{\mathbf{o}}\mathbf{o}$ objects would weigh the same "because they weighed the same before", or initially concluding that the stuff in the steel and aluminum objects was the same in heaviness, but then predicting that the steel object in the $\underline{\mathbf{o}}\mathbf{o}$ pair would be heavier). Others were more consistent, but erroneously concluded from the example that steel and aluminum are the same in heaviness. All of these errors reflect their failure to use *heavy for size* rather than *heavy* consistently in generalizations about the weights of different materials.

Older children were more succesful in the steel/aluminum preliminaries: 61% of the 5-year-olds, 75% of the 6–7-year-olds, and 83% of the 8–9-year-olds answered these questions correctly. Children could be correct on these questions, however, without fully understanding density. This is because many of them judged the steel object to be heavier than the aluminum object in the first preliminary pair ($\mathrm{O}_{=}\mathbf{o}$). Although they were shown that the two objects weighed the same, some may still have judged that the steel was a heavier kind of stuff simply because it had felt heavier. If this interpretation is correct, then children should have more difficulty answering the steel/aluminum forced choice problems than the steel/aluminum preliminaries (prediction confirmed for 3–7-year-olds, see next section). Children should also have more difficulty answering the preliminary questions about the wax and clay balls because fewer children thought that the clay ball in the $\mathrm{O}_{=}\mathbf{o}$ pair felt heavier. In fact, the wax and clay preliminaries were more difficult: none of the 3–4-year-olds, only 20% of the 5–7-year-olds and 50% of the 8–9-year-olds answered these questions correctly.

The forced choice problems

Figure 5 shows how frequently children judged the same size pieces would weigh the same, both for the steel/aluminum and wax/clay problems. Piaget's and Inhelder's finding is replicated. Size errors were very frequent among the 3–7-year-olds, clearly more frequent than the 33% size errors that would be expected if children were guessing ($p < .05$, Binomial, 2-tailed, for each of

Figure 5. *Percent of size errors as a function of age and problem type on the modified Piagetian wax and clay ball task.*



the six replications). In contrast, by ages 8 and 9, size errors rarely occurred for the steel/aluminum comparisons and the postchallenge wax/clay problems. For the younger children, the steel/aluminum and wax/clay problems were of comparable difficulty. However, the 8–9-year-olds did better with the steel and aluminum problems; their performance on the wax and clay problems only approached that of the steel and aluminum problems after their wrong answers on the wax and clay problems were challenged. The wax and clay problems may have been more difficult for them because they had to make the inference about which material was denser entirely on the basis of the preliminaries. Most 8–9-year-olds had already inferred steel was denser than aluminum prior to these problems. Younger children had difficulty with both because they did not yet have a distinct concept of density. Because of the added difficulties on the wax/clay problems, pattern analyses were done only for the steel and aluminum problems.

Most children showed one of four systematic patterns of responding on the steel/aluminum problems (Table 10). The first pattern revealed a correct understanding of the density differences of the materials: children picked a larger aluminum piece to match a target steel piece in weight and picked a smaller steel piece to match a target piece of aluminum (correct patterns). This pattern virtually never occurred among the 3- and 4-year-olds, first makes its appearance among the 5–7-year-olds and is the modal pattern among the 8–9-year-olds. Significantly, the 3–7-year-olds find the forced choice problems more difficult than the preliminaries (Sign test, $p < .001$), while the 8–9-year-olds do well on both. These results support our findings in Section 2 that it is primarily the 8–9-year-olds who clearly understand density.

Two patterns revealed children's reliance on size as a predictor of weight: some children consistently chose an object the same size as the target (size patterns); others sometimes chose on the basis of size, and sometimes chose correctly (correct/size patterns). These patterns are most frequent among the 3–4-year-olds, and decline in frequency with increasing age. Significantly, correct/size patterns are just as frequent as size patterns.

Finally, some children may not have based their judgments on size. One group seemed to choose objects on the basis of their position in an array: always choosing the object in the middle, or always choosing the end object. Others followed no systematic pattern, and were categorized as "other". These position and other patterns were also quite frequent among the 3–7-

Table 10. *Steel/aluminum and wax/clay forced problems: Patterns of response as a function of age*

Pattern	Age				
	3 (N= 18)	4 (N = 18)	5 (N = 18)	6–7 (N = 12)	8–9 (N = 12)
Correct	–	6%	38%	42%	83%
Correct/size	33%	28%	11%	8%	17%
Size	22%	38%	17%	25%	–
Position	28%	22%	17%	7%	–
Other	17%	6%	17%	8%	–

year-old children. These children thus give no evidence of using size as a predictor of weight.

Why do more children use size as a basis for their judgments on the forced choice problems than use size in their initial predictions in the preliminaries? After all, the experiences that intervene all provide evidence that size is *not* a predictor of weight in these problems (i.e., they are shown that $\mathrm{O}_{=}\mathbf{o}$ items have the same weight and $\underline{\mathbf{o}}\mathrm{o}$ items have different weights). One possible explanation is that in the preliminary question children had the option of lifting the objects. For the most part, children either expected the steel object to be heavier or were uncertain enough that they wanted to lift the objects. However, in the forced choice problems they were not allowed to say that the steel object was always heavier: they were told to find the piece that equaled the steel in weight. Further, they were not allowed to find the answer empirically by lifting and comparing the different objects in felt weight. The prevalence of size and correct/size patterns shows that they did know that size is a predictor of weight. The fact that they use this knowledge only when forced to argues that they at best consider size to be a rough indicator of weight.

Many children also gave evidence of having generalizations relating material kinds and weight. Some children initially predicted that the small piece of steel would be heavier than the large aluminum piece. The majority of 5–9-year-old children both stated, in the steel/aluminum preliminaries, that the steel was a heavier kind of stuff and predicted that a piece of steel would be heavier than a same size piece of aluminum. And many children (approximately half of the 4–5-year-olds and virtually all the 6–9-year-olds) were able to explain why a ball of clay is heavier than a same size ball of wax by appealing to the materials the objects are made of (e.g., "they are made of different materials", "one is clay, the other is wax", "clay is heavier", "clay is a heavier kind of stuff", etc.). Thus, from age 4, children could frequently articulate generalizations like "steel is heavier than aluminum". However, only the 8–9-year-olds consistently interpreted the "heavier" as "denser" as indicated by their pattern of predictions on the forced choice problems.

There is less evidence that the 3-year-olds had generalizations like "steel is heavier". They were usually inconsistent in their answers to the steel/aluminum preliminaries, and only 17% of them could explain why the clay ball was heavier than the same size wax ball by appealing to the different materials the objects were made of. Significantly, in the absence of this knowledge, these children did not expect size and weight to perfectly covary. They too expected only a rough correlation as indicated by their preference for feeling the objects before judging their weights and the paucity of systematic size patterns.

In conclusion, some young children may rely exclusively on felt weight cues in making weight judgments. Others have formed crude empirical generalizations linking size and weight and kind of material to weight. Initially, however, these two generalizations are uncoordinated. Only the older children begin to coordinate these generalizations in a concept of density.

4. Matter and material kinds

Piaget and Inhelder (1974) showed that if a ball of clay is flattened or divided into little pieces, young children maintain that the amount of clay has changed. This finding raises the question of what young children think clay is. Do they realize that clay refers to the kind of material in the ball or pancake? If so, what do they consider the defining properties of clay? Or, do they think that clay is a type of object from which other objects can be fashioned? The data presented in Sections 2 and 3 also raise questions as to what children understand about material kinds. The density we are studying is a property of material kinds, while weight is a property of objects. Perhaps one source of young children's extraordinary difficulties in understanding the sense in which steel is heavier than aluminum is problems with concepts such as *steel* and *aluminum*. In Section 4 we explore whether young children have a concept of kind of matter which is distinct from kind of object. We wish to know when children have a notion of material kinds clearly available, and what properties they take to be characteristic of material kinds.

Students of ontology have proposed the "universal grinder test" to distinguish kinds of objects and kinds of stuff. Materials, if run through a grinder that can cut them into arbitrarily small pieces, still retain their identity—gold in, gold out; sand in, sand out; water in, water out. Not so for kinds of objects—a table in, wood out; a window in, shattered glass out, and so on. Further, different properties of any given object are due to its participation in the two levels of classification. You eat on a table because it's a table; you can burn it because it is made of wood. We want to know to what extent young children make this distinction between kinds of materials and kinds of objects. If they do, we wish to know what properties characterize material kinds. In particular, do they define material kinds in terms of properties of large scale chunks or in terms of properties of underlying constituents? Is weight (or density) seen as an important property of material kinds?

Method

Subjects

Thirty-two subjects (8 4-year-olds, 5-year-olds, 6–7-year-olds, and 8–9-year-olds) participated in a structured interview about material kinds. These subjects were chosen randomly from the 78 subjects who had participated in the tasks already discussed. In addition, 24 subjects (the same 16 4–5-year-olds, and another 8 6–9-year-olds selected from the original sample) participated in a structured interview about weight .

The interview about material kinds

The material kind interview began with the question "Do you know what kinds of stuff different things are made of?" A plastic cup and a glass cup were then produced, and the child was asked of each: "What kind of stuff is this made of?" If the child answered correctly, he or she was asked what were the differences between plastic and glass. The interview continued by probing what kinds of stuff tables, bicycles, dogs, clouds, and the sun are made of, and probing the differences between wood and metal. Then the child was asked whether shadows are made out of some kind of stuff, and what the difference between shadows and tables is, in terms of their being made of some kind of stuff.

The second part of the interview was inspired by the universal grinder test. Four different objects (a paper cup, a rubber balloon, a wooden airplane, and a metal spoon) were presented to the child, and the child was asked what they were and what they were made of. Then the objects were cut into small pieces as the child watched[8] and the child was asked whether the cut-up object was still the same kind of object and whether it was still the same kind of stuff (e.g., "Is it still a cup? Is it still paper?"). For two of the problems, the kind of object question came first and for the other two the material kind question came first.

The third and fourth sections of the interview probed the child's understanding that weight (density) is a fundamental property of material kinds. First, the child was shown two metal cylinders of identical size and shape. In one case the cylinders were both brass, but one was shiny and the other corroded. In the other case, one cylinder was aluminum and one steel, but they had been identically painted with metallic paint. For both pairs (presented in

[8]In the case of the metal spoon, small metal pieces were produced and the child was told that they were the result of cutting up a spoon just like the whole one.

counter-balanced order across subjects) the child was allowed to pick up the objects and was asked whether or not the two objects were made of the same kind of metal. If children said they were made of the same kind of metal, they were asked if it was *possible* that the two could be different metals. If children said they were made of different metals, they were asked if it was *possible* that they were made of the same kind of metal.

The fourth task was adapted from Bovet, Domahidy-Dami, & Sinclair (1982). The child was shown a block of clay and a block of playdough which were the same size and he/she established by lifting the blocks that the clay block was heavier. The blocks were divided in two several times. Each time the child was asked if one block was heavier or whether they were both the same weight. However, the child was not allowed to pick up the blocks. If children maintained that the clay was heavier until this point, they were asked whether this would always be the case, no matter how small the two pieces became. All children were then shown tiny pieces of clay and playdough (pieces about the size of BBs) and were asked whether one piece was heavier, or whether they both weighed the same.

The weight interview

The weight interview included a series of questions designed to probe whether the child thought that all objects must weigh something. The child was first given a piece of styrofoam and was asked: "Does this piece of styrofoam weigh a lot, a tiny, tiny bit, or nothing at all?" Then the child was shown a tiny piece of playdough and was asked if it weighed anything at all. The piece was added to a much larger piece of playdough and the child was asked if the ball of playdough was heavier after the tiny piece had been added or whether it still weighed the same. Finally, similar questions were asked about a grain of rice and a pile of rice with one grain added.

Results

Material kind interview

Table and cups

The child's first task was to provide some examples of the kinds of stuff things are made of. At all ages, when children responded at all, their answers were relevant examples of material kinds. Younger children were less likely to respond than older children (50% were able to respond among the 4–5-year-olds vs. 87% for the 8–9-year-olds) and gave fewer examples (one or two examples among the 4-year-olds vs. five or six examples among the oldest children). Nonetheless, further questioning revealed that every child knew some words for material kinds. They all knew that the cups were made of plastic and glass. And, from age 5 on up, all children were able to state relevant differences between plastic and glass (as were over half the 4-year-olds). For example, glass is heavier, harder, more breakable than plastic. Children were less able to articulate the differences between wood and metal. However, when children were able to articulate a difference, it was almost always a difference relevant for material kinds.

The universal grinder test

There were virtually no errors on the grinder test. At all ages the children knew that the cut-up objects were no longer the same kinds of objects but that they were still the same material kinds. Justifications changed with age. Half of the younger children (4–7-year-olds) mentioned only perceptible properties of the cut-up pieces in explaining why it was still the same material kind (i.e., "it is still shiny" or "it is still sharp"). In contrast, older children and the rest of the younger children responded in a different manner. They explicitly stated principles such as, "cutting does not affect material", "it's still paper because the cup was *made of paper*" (child's emphasis), or were groping towards expressing such principles by saying "you just cut it up, it's still the same".

Dogs and bicycles

Children's ability to distinguish between kinds of objects and material kinds was further examined by asking them what kinds of stuff bicycles and dogs are made of. Children who are unclear about this distinction might give parts rather than material kinds as answers. Only two children responded with a part of a bicycle (wheels); the rest referred to material kinds such as metal and plastic. Similarly, only six children (spread over all four ages) responded with external body parts of a dog. The rest of the responses were the same in flavor as the songwriter's "muscle and blood and skin and bone ..."

Interim summary

All of the children in this sample knew several material kind words, and virtually all distinguished between objects and material kinds in three ways: they knew that cutting destroyed objecthood while not affecting material kind, they knew relevant properties of specific material kinds, and they were not misled into giving parts of objects when asked what kind of stuff they were made of. However, these achievements by no means guarantee that young children have the same conception of material kind as older children or adults. The changes in children's justifications on the universal grinder test suggest one important difference: younger children may define a material kind in terms of the perceptible properties that they use to identify it; older children may define a material kind in terms of its being a fundamental constituent of an object and realize that many of its perceptible properties aren't essential characteristics of that kind. Analysis of the rest of the material kind interview provides further evidence about the ways in which children's concept of material kinds is changing during these years.

Shadows

The questions about whether shadows are made of some kind of stuff, in the same sense that tables are, bear on how children conceptualize "made out of" as well as how they think about "kind of stuff." Most young children seem to interpret the expression "is made out of" to mean "is constructed from" rather than "is constituted of" or "contains that kind of stuff at every point." One-third of the 4–7-year-olds said that shadows were made of some kind of stuff, just like tables, explaining that the object which casts a shadow and/or the sun is the kind of stuff they are made of. For example, "they are made of you" or "they are made of you and the sun." Another third of the youngest children appeared to make the same misinterpretation, but their answers were more ambiguous. For example, when asked "Are shadows made of some kind of stuff?", they responded "yes, people stand in front to make a shadow" or "no, they are made out of people." None of these children could offer a relevant difference between shadows and tables. Finally, the oldest children in the sample and a third of the younger children as well gave clear evidence of conceptualizing materials as underlying constituents. Some of them made a clear distinction between material and nonmaterial objects. They said that shadows, unlike tables, are not made of some kind of stuff, and supported their answer with either a relevant justification or a relevant difference between shadows and tables (e.g., "shadows are only a reflection of your body", "you can feel a table, but you can't feel a shadow"). Other children asserted that shadows, like tables, were made of something, but revealed in their justifications that they were thinking of an underlying constituent. For example, "they are made of shade and cold atoms; that's why shadows are dark and cool" or "they are made out of darkness, the same as is night".

If children do not think of materials as underlying constituents of objects, then the perceptual properties of the materials should be more essential in their definition. Direct support for this hypothesis comes from an analysis of the relation between children's patterns on the shadows task and their type

Table 11. *Relation between children's pattern on the shadow questions and their type of justification on the grinder task*

Pattern on shadow questions	Type of justification on the grinder task			
	Appeal only to perceptual properties or don't know ($N = 13$)		Appeal to general principles based on material: "Cutting does not affect material"; "It's made of X" ($N = 19$)	
Interpret "made of some kind of stuff" as "is constructed from some kind of stuff" ($N = 6$)	83%	(5)	17%	(1)
Ambiguous with respect to constructed from/is constituted of distinction ($N = 10$)	50%	(5)	50%	(5)
Show understanding of notion of material object (i.e., an object which is constituted of some kind of stuff) ($N = 16$)	18%	(3)	82%	(13)

of justification on the grinder task. Table 11 shows that 83% of the children who clearly interpreted "made of" as "constructed from" in the shadow questions only appealed to perceptual properties of the materials in their justifications on the universal grinder task. In contrast, 70% of the children who interpreted "made of" as "constituted of" in the shadow questions also appealed to general principles like "cutting does not affect materials" in the grinder task ($\chi^2 = 8.3$, df $= 2$, $p < .05$).

Painted steel and aluminum/shiny and corroded brass

In this task, we explored whether children thought that weight differences were more important than differences in surface appearance in distinguishing material kinds. Children were shown two pieces of steel and aluminum which were the same in size and color but were different in weight; they were also shown two pieces of brass which were the same in size and weight but were different in surface appearance. In each case, they were asked whether they thought the objects were made of the same kind of stuff. Depending on their answer they were then asked whether it was at least possible that they were made of the same kind of stuff or possible that they were made of a different kind of stuff.

There were three main patterns of response (see Table 12). Most of the older children said that the painted steel and aluminum cylinders must be made of different kinds of stuff, because they differed in weight. At the same

Table 12. *Painted steel and aluminum/shiny and corroded brass task: Pattern of response as a function of age*

Pattern	Age			
	4 ($N = 8$)	5 ($N = 8$)	6–7 ($N = 8$)	8–9 ($N = 8$)
Painted steel and aluminum cylinders must be made of different materials, but shiny and corroded brass cylinders could be made of the same material	–	25%	25%	75%
Both painted steel and aluminum cylinders and shiny and corroded brass cylinders must be made of different materials	25%	38%	38%	25%
Shiny and corroded brass cylinders must be made of different materials, but painted steel and aluminum cylinders could be made of the same material	12%	25%	12%	–
Both painted steel and aluminum cylinders and shiny and corroded brass cylinders could be made of the same material	62%	12%	25%	–

time they thought that it was possible that the shiny and corroded brass could be made of the same kind of stuff, because they were the same in weight. These children clearly consider weight differences more important than surface appearance differences in distinguishing material kinds. A different pattern was shown by many of the 5–7-year-olds. They said that both the painted steel and aluminum and the shiny and corroded brass must be made of different kinds of stuff. Many of these children surprisingly justified both choices in terms of weight differences: they correctly said that the steel piece was heavier than the aluminum and incorrectly said that the corroded brass piece was heavier than the shiny piece. These erroneous judgments may reflect their belief that dark things are heavier (a belief many of them explicitly stated). Although these children take weight to be important in distinguishing material kinds, they also expect a correlation between weight and surface appearance. A third pattern was typically shown by the youngest children. These children said that both the painted steel and aluminum and shiny and corroded brass could be made of the same kinds of stuff. These children then regard neither weight nor surface appearance as central to determining material kinds. Further work with other properties would be necessary to determine whether these children consider some other properties (e.g., texture cues) essential in distinguishing material kinds. Finally, a fourth possible pattern (judging that the shiny and corroded brass must be made of different kinds of stuff, but that the painted steel and aluminum could be made of the same kind of stuff) rarely occurred. When it did occur, it was among the younger children. This pattern is consistent with their considering surface appearance differences to be more important than weight differences in distinguishing material kinds.

Halving of playdough and clay

Children's performance on the painted steel and aluminum task shows that from age 5 on up they regard weight differences as important for distinguishing among material kinds when questioned about large-scale objects. In this task, we present children with two same size pieces of playdough and clay which differ in weight. We ask whether children expect that the weight difference between clay and playdough will be preserved as the objects are successively halved until they are very tiny (i.e., the size of BBs).

There were three main patterns of response to the questions about the halving of playdough and clay. Half of the oldest children judged that the clay ball always weighed more than the playdough ball even in the case of the two tiny pieces (Table 13). These children typically explained that the objects were made of different materials and were still made of those different materials even when very small (e.g., "It's still clay. It's still made out of the same thing, no matter how small you make it.") More common among the 5–7-year-old children was to maintain that the clay ball was heavier than the playdough for several of the halvings, but then, when the two balls were

Table 13. *Halving of playdough and clay: Patterns of response as a function of age*

Pattern	Age			
	4 ($N = 8$)	5 ($N = 8$)	6–7 ($N = 8$)	8–9 ($N = 8$)
No maintenance of weight differences	62%	25%	38%	13%
Maintain weight differences until balls are very small	38%	62%	38%	38%
Adult pattern: clay ball will always weigh more	–	13%	25%	50%

sufficiently small, to switch to the judgment that the two weighed the same "because they don't weigh anything at all" or "because they are both light". Finally, the most common pattern for the youngest children was to switch back and forth between the judgment that the clay ball was heavier, the playdough ball was heavier, or the two were the same in weight, in an apparently unprincipled manner. Children with this pattern typically referred to the object's size, height, and appearance (e.g., "it looks heavier") in their justifications, when they provided justifications.

Overall, few children took weight to be a fundamental property of small-scale objects. Why should there be this discrepancy between children's beliefs about the properties of large scale and small scale objects? One hypothesis is that felt weight is at the core of the young child's weight concept. It is because the tiniest pieces of playdough and clay feel like they weigh nothing at all that the younger child believes that they do not weigh anything. In contrast, the core of older children's weight concept may be that it is a fundamental property of material objects. Thus, for them as long as an object has matter, it must have weight. The last set of tasks we report in this section directly tests this hypothesis by probing whether children believe light objects weigh anything at all.

The weight interview

The majority of 4–6-year-olds (75%) thought that the styrofoam ball weighed nothing at all. They justified their answers by saying that it did not feel like it weighed anything or by saying that the substance it was made of was light. In contrast, the majority of 8–9-year-olds (75%) said that the styrofoam ball weighed a tiny, tiny bit, because "everything has to weigh something."

The patterns on the playdough and rice questions also confirm that young children think of weight as felt weight while the older children are beginning to think of weight as a property of matter. Many of the 4–6-year-olds (45%) said that a small piece of playdough and rice did not weigh anything and would not make an object/pile heavier when they were added to it, justifying their judgments by invoking the object's felt weight or size. They all agreed, however, that adding a big chunk of playdough to the playdough ball or many grains of rice to the pile of rice *would* make a difference in their weights. Most of the other 4–6-year-olds (55%) were less consistent: they typically said that a small piece of playdough weighed something and would make a playdough ball heavier, but noted that a small piece of rice weighed nothing at all and would not make a pile heavier. In contrast, most of the older children (75%) and one 5-year-old consistently judged that both the playdough and rice weighed something and made the playdough ball/pile of rice heavier, justifying their judgment by saying "everything weighs something".

Conclusions

We conclude then that children as young as age 4 have some notion of material kind that is different from their notion of kind of object. However, there seem to be three main ways their notion of material kind is changing during the age period from 4 to 9. First, most 4–7-year-olds either consider weight and surface appearance differences to be essential in distinguishing material kinds, consider neither to be essential, or consider surface appearance differences to be more essential than weight differences in distinguishing material kinds. In contrast, most 8–9-year-olds consider weight differences to be more important than surface appearance differences in distinguishing material kinds. Second, most 4–7-year-olds interpret "made out of" as "constructed from" and rely on the perceptual properties of large scale chunks of a material in judging its identity. In contrast, most 8–9-year-olds clearly interpret "made out of" to mean "is a constituent of" and reason that materials must remain the same kind of stuff even when they are ground up. Finally, felt weight is at the core of most 4–7-year-olds' concept of weight. In contrast, many 8–9-year-olds consider weight to be a necessary property of matter and are beginning to appreciate that density differences are preserved for even small pieces of different kinds of stuff. These changes do not all occur at precisely the same time in an individual child, but they mutually reinforce one another and lead to significantly new ways of thinking about material kinds.

What seems to be developing during these years, then, is a new level of description—a micro-level relative to the level of objects. For a table to be constituted of wood, as opposed to merely constructed out of wood pieces, means that it is wood at every point. It is not necessary that the child think of the points in their limit, invisible to the naked eye, or that they be spatially conceived so as to be amenable to the schemas of compression and expansion. Indeed, we do not believe children have a particulate theory of matter at this time. But it is necessary that small pieces be conceivable as tiny chunks of wood or clay that maintain their identity and properties such as relative heaviness for size.

Discussion

Our goal in this case study was to examine two claims: (1) that individual concepts undergo differentiation during child development; and (2) that young children's concepts are embedded in theory-like structures. Two putative cases of differentiation in children were considered: the differentiation of the concepts of size and weight and the differentiation of the concepts of weight and density.

Our results show that differentiation at the level of individual concepts does occur in young children and is like conceptual differentiation in scientists in two important respects: (a) the undifferentiated parent concept can be analyzed as containing distinct components of the descendant concepts united in one conceptual unit, and (b) the undifferentiated parent concept occurs in a different theoretical structure from the differentiated descendant concepts. We found that children conflate the components *heavy* and *heavy for size* in one weight concept and that they develop distinct weight and density concepts as they are reconceptualizing their concept of material kinds. In concluding, we will first summarize our argument that children's concepts of weight and density are differentiating and are embedded in theory-like structures. We will then consider the wider implications of our findings for other work in cognitive development.

To trace the development of children's concepts of weight and density, we have simultaneously followed the development of some other closely allied concepts in children's physical theory: *size*, *object*, *matter*, and *material kind*. We argued that young children (approximately ages 3 and 4) have early theories which contain distinct concepts of weight, size and object as well as emerging concepts of material kinds (e.g., *glass* and *plastic*). The core of their weight concept is *felt weight* as indicated by their insistence that a piece of styrofoam weighs nothing at all. But there are other components of their weight concept as well. Weight is considered a physical property of objects which causally affects that object's interactions with other objects, as indicated by their belief that a heavy object will make a foam rubber bridge collapse.

Two more points can be made about children's early concept of weight: (a) the component *perceived size* is not part of their weight concept; and (b) the component *heavy for size* is not yet part of this concept. Thus, their concept of weight is differentiated from their concept of size, and the concept of density is totally absent.

There were several lines of evidence that young children have distinct size and weight concepts. First, young children could selectively focus on perceived size (ignoring felt weight) to explain certain physical phenomena (i.e., which blocks will fit into a certain size box) and selectively focus on felt weight (ignoring perceived size) to explain other physical phenomena (i.e., which blocks will make a foam rubber bridge collapse). Second, young children had learned distinct words for size and weight. Third, some young children might not have even a crude expectation that size is a predictor of the heaviness of objects, since they failed to use size as a predictor in the wax and clay ball task. Fourth, those who had noted a relation between size and weight conceptualized the relation as one of a crude empirical correlation.

When given the choice, they preferred to judge weights by lifting objects. Finally, the notions of size and weight figure in separate generalizations in their theory. Some children had formed a generalization that steel is heavy and totally ignored size in predicting weight. The picture that emerges, therefore, is that the concepts of size and weight have separate sensory cores, play different roles in children's theories, and are even beginning to be interrelated.

Children's early concept of weight contains no traces of a concept of density. This is indicated by the fact that many of the 3- and 4-year-old children judged exclusively on the basis of felt weight even on the nonverbal density task where there had been pretraining to make the notion *heavy for size* more salient.

Between ages 5 and 7, children's concept of weight becomes modified, although the core remains *felt weight*. The component *heavy for size* is now added to the cluster of components characterizing children's notion of weight, and it is at this point that they can be considered to have an undifferentiated weight/density concept. Significantly, this change in the child's concept of weight occurs during the same age period children are making changes in their concept of material kinds. They are increasingly able to articulate generalizations relating weight and material kinds, and they now regard weight differences as important in distinguishing whether objects are made of the same material kind.

The main evidence that children have two distinct senses of *heavy* available to them (i.e., *heavy*, and *heavy for size*) comes from the fact that they made both kinds of judgments in the weight and/or density tasks. Evidence that they conflate the two components in one concept comes from an analysis of how the components function in the children's overall theory. These children gave no evidence of using the components in separate generalizations. Rather they used both components in generalizations about the weight of material kinds, especially when pretraining had made the *heavy for size* component more salient (nonverbal density tasks). Although they typically used only the component of absolute weight on the verbal weight task, there had been no pretraining to make the component *heavy for size* more salient and the structure of the task (involving the direct comparison of two objects) makes absolute weight differences more salient. These same factors lead to more absolute weight judgments on the verbal density task as well. Weight errors on the verbal density task are particularly significant because children could visually identify the different materials and had been shown in pretraining that a piece of brass was heavier than a same size piece of aluminum. These problems should be easier than the nonverbal density tasks for children who have a concept of density, since they do not need to make judgments solely on the basis of felt weight adjusted for size. In fact, the 3–7-year-olds found these problems no easier; they typically judged only on the basis of absolute weight and kind of stuff. We are arguing, then, that when both components *heavy* and *heavy for size* are made salient, children will use both components indiscriminately in generalizations about the weights of material kinds and in generalizations about the weights of objects; otherwise, they will rely primarily on absolute weight. In keeping with this expectation, we found that density intrusions on the verbal weight task did occur when the child had had a density task preceding it in the session.

The theoretical context not only is necessary to decide issues of differentiation, but also helps one understand why children unite these two quite different components in one concept. Weight is a concept which has a built-in comparative structure: it is a dimension along which objects vary and can be compared. For young children, this dimension is *felt weight*. Further, judgments of felt weight are always made relative to some standard of comparison: an object may be judged heavy relative to another specific object, heavy relative to other objects of its type, heavy relative to other objects of comparable size, or even heavy relative to the person doing the lifting. Seen in this light the notion *heavy for size* is by no means anomalous as part of the child's weight concept. It simply fits into the comparative structure of this concept and expresses one relevant relativization. Further, the larger theoretical context prevents the child from clearly seeing that a different sense of weight is called for in generalizations about material kinds. Although the young child does have some concept of material kind, it is a decidedly nonadult notion. For them, material kinds are defined in terms of properties that characterize large-scale chunks of that stuff (e.g., being heavy or light, rough or smooth, brittle or hard). They do not think of material kinds as underlying constituents of the object at a micro-level. The relation they establish between weight and kind of stuff is of the type: "steel objects are heavy, wood objects are light" rather than "steel is a heavy material; wood is a light material". These generalizations have the same status to the child as other generalizations about the weights of objects (e.g., "tables are heavy, balloons are light"). And because all these generalizations are really about large-scale objects rather than materials, it makes sense that they all call for the same concept of weight.

How then do children come to realize that *heavy for size* is an important physical variable, characteristic of material kinds, which needs to be distinguished from weight? The motivation might be empirical. Children might notice that steel objects, for example, are always heavier for their size but not always heavy. They might also notice that some physical effects depend on weight and not on weight relativized to size (e.g., making a foam rubber bridge collapse). However, in the child's everyday experience, objects made of some kinds of stuff (such as styrofoam) are always light, in the absolute sense, while others (such as steel) are always heavy. Thus, the child's experience alone may not be sufficient to convince him that objects made of certain kinds of stuff are heavy for their size rather than absolutely heavy. Changes within the child's conception of matter and material kinds may also be important in supporting the differentiation between weight and density. The child develops a micro-level of description in which weight is seen as a fundamental property of matter and density a fundamental property of material kinds. This micro-level of description allows the child to see clearly the relation between the weight of large-scale objects and the weights of the different kinds of tiny pieces of matter of which they are composed.

Some of the 8- and 9-year-olds in our study had reached the point where they not only had distinct density and weight concepts but also had reconceptualized weight as a property of matter and reconceptualized material kinds as the underlying constituents of objects. Such children systematically relativized weight to size in the density tasks, asserted that everything must weigh something, no matter how small it is, and judged that a small piece of clay is still heavier than a piece of playdough the same size. This reconceptualization of weight and material kinds at a micro-level of description is a first important step in allowing the child to see the difference between weight and density. But it is not the last word. Children at this time probably still have not developed the idea of a standard unit of volume and hence conceptualize density as heaviness for size rather than as weight per unit volume. Futhermore, lacking such standard units, they would not yet attempt to calculate densities numerically or realize that there is a unique number which defines the density of a substance under standard conditions. Finally, they probably have not conceived of explaining the density differences of solids in atomistic terms, since Piaget himself has shown this to be a much later development.

We agree, then, with Piaget's and Inhelder's conclusions that *weight* and *density* undergo differentiation during development. Like them we have also placed this differentiation in the context of a developing theory of matter. However, there are important differences in the two accounts which have implications beyond this particular case study. Piaget and Inhelder place the development of weight/density concepts not only in the context of theory change but also in the context of more structural changes in cognitive capacity. The young child is seen as not being capable of representing true concepts until the attainment of concrete operations at around ages 7 and 8 and as not being capable of representing theories until the attainment of formal operations in adolescence. Consistent with their claims, Piaget and Inhelder describe young children as initially having a size/weight concept which is a diffuse, phenomenalistic whole. Furthermore, they see the emergence of the child's first physical theory (atomism) as a relatively late development.

In contrast, we assume that even 3-year-old children can represent true concepts and that an undifferentiated concept is not different in kind from a differentiated concept. Consistent with our claims, we found that children's undifferentiated weight/density concept consists of distinct components (e.g., *heavy*, *heavy for size*) which can be selectively used in different contexts. It is hard to imagine what a diffuse syncretic weight/density concept would be. There is no direct sensory basis for the component *heavy for size*; so it would not figure in the kind of immediate, wholistic perception of objects described by Kemler (1983). Moreover, the youngest children in our sample used a concept of weight which was uncontaminated by *size* or *heavy for size*. Thus, the developmental progression in their concept of weight is from weight to undifferentiated weight/density to differentiated weight and density concepts. In traditional developmental accounts (e.g., Bower, 1974; Piaget & Inhelder, 1974; Werner, 1948), the most primitive state is the undifferentiated one.

We also assume that young children can represent theories. Consistent with our claims, even young children do more than conceptualize weight as *felt weight*. They also conceptualize weight as a physical variable—that is, as an attribute of an object which affects its interaction with other objects and which can be crudely predicted from knowledge of the object's characteristics. And they have some concepts of matter and material kind. By age 8, many children have deepened their theoretical understanding of weight by reconceptualizing it as a fundamental property of matter. They now have principled generalizations that the weight of an object is a function of both the amount of matter and the kind of matter in that object. Significantly, this deepening of their physical theory as a matter theory occurs in the early elementary school years well before their matter theory takes a clear atomistic form. Finally, we found that children at all ages were able to verbalize their understandings as well as act on them. Indeed, one of the striking findings of our study was the comparable difficulty of the verbal and nonverbal size, weight, and density tasks. Hence, children's early theories cannot be dismissed as qualitatively different theories of action, as has been suggested recently by Karmiloff-Smith and Inhelder (1974/75). Thus, we conclude that the young child can be more of a theorist than has been commonly supposed in the developmental literature.

Our account of conceptual development in terms of theory change not only differs from the Piagetian tradition but also from the information processing approach in developmental psychology. Information processing theorists have been particularly interested in developing models of how children solve certain problems and how children's approaches to these problems change with age. In these models, the child of a given age is characterized as holding certain rules; with increasing age, the child comes to hold rules of increasing adequacy and complexity. No attempt is made, however, to characterize the concepts which are used in these rules or the conceptual structures which guide the child's selection of rules to use in a particular problem situation.

A classic example of the information processing approach is provided by Siegler's analyses of children's performance on balance scale problems (Siegler, 1976). In these problems, the child is shown an arrangement of weights at varying distances from the fulcrum and is asked to predict whether the arrangement will balance when the experimenter releases it. Siegler argues young children simply check if the weights are the same on each side. Slightly older children take both weight and distance from the fulcrum into account but do not know how to coordinate them. Finally, the oldest children check whether the cross-product of weight and distance is the same on both sides. Although Siegler provides a clear account of changes in children's solutions to these problems, he does not attempt to investigate what children's concept of weight is, whether this concept of weight is changing, or whether new conceptual units are created in the course of understanding the balance scale problems. It is implied that developmental change involves forming increasingly complex rules, using the same basic stock of concepts. Our own work suggests that the restructurings which occur in development can be of a more radical type than that envisioned by information processing theorists. Like the reorganizations which occur in theory changes in the history of science, some developmental reorganizations also involve changes in the conceptual units. We have already documented important changes which occur in the child's concepts of weight and material kinds between the ages of 3 and 7. We suspect that other fundamental restructurings may occur in the concepts the child applies to understanding the balance scale (e.g., *force*, see Piaget, 1974).

In conclusion, we have argued that Kuhn's account of conceptual differentiation in the history of science is applicable to conceptual differentiation in children. Further, this account which analyses concepts as they participate in theories, provides answers to two fundamental descriptive questions: (1) how to trace descent between concepts, and (2) how to explain the sense in which the parent concept is a precursor of the descendants. Clearly, there are important differences between the theorizing of scientists and children. Scientists set out to self-consciously develop theories; they plan experiments to test theories; and they even have philosophical positions about the nature of theories. Although young children are aware of the content of their theories, they are not as aware that they are engaged in a general process of theorizing. An important question for further research, then, is to understand how this lack of metaconceptual awareness about theories affects the actual process of theorizing in children. Finally, we would stress that our work has addressed only the issue of providing clearer *descriptions* of conceptual differentiation. As yet, we have no theories about the underlying *mechanisms* of conceptual differentiation and do not know whether common mechanisms underlie conceptual differentiation in children and scientists. Our hope, however, is that greater descriptive clarity will aid our search for underlying mechanisms.

Appendix 1: Stimuli

Nonverbal size and weight tasks

All the items were cubes

	Dimensions (in.)	Weight (g)
Practice items		
Painted small styrofoam	$2 \times 2 \times 2$	5
Painted small aluminum	$2 \times 2 \times 2$	320
Painted large styrofoam	$4 \times 4 \times 4$	40
Painted large plexiglass	$4 \times 4 \times 4$	1,100
Test items		
Small plexiglass	$1\frac{1}{2} \times 1\frac{1}{2} \times 1\frac{1}{2}$	60
Small aluminum	$1\frac{1}{4} \times 1\frac{1}{4} \times 1\frac{1}{4}$	80
Small brass	$1\frac{1}{4} \times 1\frac{1}{4} \times 1\frac{1}{4}$	260
Small steel	$1\frac{1}{2} \times 1\frac{1}{2} \times 1\frac{1}{2}$	440
Large styrofoam	$4 \times 4 \times 4$	42
Large balsa wood	$4 \times 4 \times 4$	127
Large plexiglass	$3\frac{1}{2} \times 3\frac{1}{2} \times 3\frac{1}{2}$	760
Large birch wood	$4 \times 4 \times 4$	790

Nonverbal density task

	Dimensions (in.)	Weight (g)
Practice items		
Uncovered steel cylinder	diam: l/height: 1½	151
Uncovered aluminum cylinder	diam: l/height: 1½	53
Uncovered steel cylinder	diam: l/height: 4	400
Uncovered aluminum cylinder	diam: l/height: 4	139
Covered steel cylinder	diam: l/height: 3	304
Covered aluminum cylinder	diam: l/height: 3	108
Covered steel cylinder	diam: l/height: 4	401
Covered aluminum cylinder	diam: l/height: 4	143
Covered steel cube	½ × ½ × ½	57
Covered aluminum cube	½ × ½ × ½	20
Covered steel block	3 × 1¼ × 1½	729
Covered aluminum block	3 × 1¼ × 1½	262
Test items		
All items were covered blocks		
1 steel	3 × ⅛ × 1½	72
2 steel	3 × ¼ × 1½	148
3 steel	3 × ¾ × 1½	441
4 steel	3 × 1¼ × 1½	715
5 steel	3 × 1½ × 1½	850
1 aluminum	3 × ⅛ ×1½	21
2 aluminum	3 × ¼ × 1½	54
3 aluminum	3 × ¾ × 1½	164
4 aluminum	3 × 1¼ × 1½	263
5 aluminum	3 × 1½ × 1½	300

Verbal size, weight and density tasks

For reasons unrelated to these studies, the cylinders were bored. Weights were not always proportional to the heights because the size of the bore (¼ in. diameter on the average) was not perfectly constant.

	Diameter/height (in.)		Weight (g)
Wood cylinders			
1 W	1	½	5
2 W	1	¾	6
3 W	1	1	8
4_a W	1	2	16
4_b W	1¼	2	27
5 W	1	4	53
6 W	1¼	6	54
7 W	1¼	7	55

Aluminum cylinders			
0 Al	1	1/4	9
1 Al	1	1/2	18
2 Al	1	3/4	26
3 Al	1	1	33
4_a Al	1	2	66
4_b Al	1 1/4	2	100
5 Al	1	4	132
6 Al	1 1/4	6	300
7 Al	1 1/4	7	350
Brass cylinders			
0 Br	1	1/4	26
1 Br	1	1/2	54
2 Br	1	3/4	77
3 Br	1	1	104
4_a Br	1	2	210
4_b Br	1 1/4	2	332
5 Br	1	4	430

Appendix 2

Adult psychophysical judgments

Pairs	Pair types	Felt weight ratios
4_a Br–4_a Al	oo	2.91
5 Al–5 W	oo	3.13
3 Br–4_b W	oO	4.07
6 Al–6 W	oo	4.75
4 Br–4 Al	oo	4.98
3 Br–4_a W	oO	5.45
6 Al–0 Br	Oo	5.50
4_a Br–5 W	oO	5.71
4_b Al–1 W	Oo	5.75
5 Al–4_a W	Oo	6.28
5 Al–0 Al	Oo	6.29
3 Br–3 W	oo	6.75
7 Al–1 Br	Oo	7.56
4 Br–2 Al	Oo	9.18
4_b Br–4_b W	oo	10.35
4_b Br–0 Br	Oo	11.08
5 Br–2 Br	Oo	11.44
7 Al–2 W	Oo	16.62
5 Br–2 W	Oo	32.08

References

Bovet, M., Domahidy-Dami, C., & Sinclair, H. (1982). Conservation and *décalages. Human Development*, *25* (3), 201–14.

Bower, T.G.R. (1974). *Development in infancy*. San Francisco: W.H. Freeman & Co.

Braine, M. & Rumain, B. (1983). Logical reasoning. In P. Mussen (Ed.), *Handbook of child psychology*, vol. 3. New York: John Wiley & Sons.

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

Carey, S. (1984). Cognitive development: the descriptive problem. In M. Gazzaniga (Ed.), *Handbook of cognitive neuroscience*. New York: Plenum Press.

Chi, M., Glaser, R., & Reese, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence*. Hillsdale, NJ: Lawrence Erlbaum.

Driver, R. & Erickson, G. (1983). Theories-in-action: some theoretical and empirical issues in the study of students' conceptual frameworks in science. *Studies in Science Education*, *10*, 37–60.

Fodor, J. (1972). Some reflections on L.S. Vygotsky's *Thought and Language. Cognition*, *1*, 83–95.

Gelman, R. & Baillargeon, R. (1983). A review of some Piagetian concepts. In P. Mussen (Ed.), *Handbook of child psychology*, vol. 3. New York: John Wiley & Sons.

Gibson, E.J. (1969). *Principles of perceptual learning and development*. New York: Appleton Century Crofts.

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

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

Kemler, D. (1983). Exploring and reexploring issues of integrality, perceptual sensitivity, and dimensional salience. *Journal of Experimental Child Psychology*, *36*, 365–379.

Kuhn, T. (1977). A function for thought experiments. In T. Kuhn, *The essential tension: selected studies in scientific tradition and change*. Chicago: University of Chicago Press.

Kuhn, T. (1982/83). Commensurability, comparability, communicability. *PSA 82*, vol. 2, (Philosophy of Science Association, 1983), 669–688.

Mandler, J. (1983). Representation. In P. Mussen (Ed.), *Handbook of child psychology*, vol. 3. New York: John Wiley & Sons.

Markman, E. & Callahan, M. (1983). An analysis of hierarchical classification. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence*, vol. 2. Hillsdale, NJ: Lawrence Erlbaum.

Novick, S. & Nussbaum, J. (1981). Pupils' understanding of the particulate nature of matter: a cross-age study. *Science Education*, *65* (2), 187–196.

Pfundt, H. (1981). The atom: the final link in the division process or the first building block? *Chimica Didactica*, *7*, 75–94. (ERIC Document Reproduction Service No. ED 229 281).

Pfundt, H. (1982). Pre-instructional conceptions about transformations of substances. *Chimica Didactica*, *8*. (ERIC Document Reproduction Service No. ED 229 235).

Piaget, J. (1974). *Understanding causality*. New York: W.W. Norton.

Piaget, J. & Inhelder, B. (1974). *The child's construction of quantities*. London: Routledge and Kegan Paul.

Quintero, A.H. (1980). *The role of semantic understanding in solving multiplication word problems*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Shultz, T. (1982). Rules of causal attribution, *Monographs of the Society for Research in Child Development*. Chicago: University of Chicago Press.

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

Suppe, F. (1974). *The structure of scientific theories*. Urbana, IL: University of Illinois Press.

Werner, H. (1948). *Comparative psychology of mental development*. Chicago: Follet.

Wiser, M. & Carey, S. (1983). When *heat* and *temperature* were one. In D. Gentner and A. Stevens (Eds.), *Mental models*. New York: Academic Press.

Résumé

On a étudie les concepts de taille, de poids, de densité, de matière et de substance chez les enfants de 3 à 9 ans. On cherche à vérifier deux propositions (1) que des concepts individuels encourent un processus de différentiation pendant le développement (2) que les concepts des jeunes enfants sont enchâssés dans des structures de type théories. Pour étudier le premier point il est nécessaire de spécifier en termes de représentations ce qu'est un concept indifférencié et dans quel sens ce concept indifférencié est apparenté à des concepts plus différenciés. La stratégie utilisée consiste à guider la recherche des preuves avec un modèle de différentiation de concepts suggéré par l'histoire des Sciences. Dans ce modèle les concepts indifférenciés tout comme les concepts différenciés peuvent être analysés en termes de propriétés componentielles, de traits ou de dimensions. La différence primordiale tient à ce qu'un concept indifférencié rassemble des composants qui seront analysés plus tard comme composants de concepts différents et que le concept indifférencié est inclu dans une structure théorique différente de celle du concept différencié. 78 enfants (18 de 3 ans, 18 de 4 ans, 18 de 5 ans, 12 de 6–7 ans et 12 de 8–9 ans) ont effectué des tâches mettant en jeu leur compréhension de la taille, du poids et de la densité. Un sous-groupe a reçu des tâches supplémentaires mettant en jeu les concepts de matière et de substance. Les données montrent que les plus jeunes enfants ont un système théorique incluant des concepts distincts de poids, de taille et de matière et forment des généralisations en relation avec ces concepts (par ex. la taille est grossièrement corrélée au poids, les objets d'acier sont lourds etc). Le centre du concept de poids est le poids senti, le concept de densité est absent du système, les substances de différents types sont définies en terme des propriétés caractérisant de gros blocs du matériel. Les enfants un peu plus âgés (5–7 ans) ont modifié leurs concepts de poids et de substance. Le concept de poids inclut alors les propriétés de *lourd* et *lourd pour la taille* (nous estimons que cela prouve un concept indifférencié de poids/densité). Ces enfants commencent à considérer les différences de poids comme importantes pour distinguer la matière dont sont faits des séries d'objets. Toutefois l'essentiel du concept de poids est encore le poids senti et les types de matériaux sont encore définis en termes de propriétés d'objets de grandes dimensions. Les enfants de 8–9 ans possèdent un système théorique dans lequel le poids et la densité sont articulés comme des concepts distincts, les types de substances sont reconceptualisés comme des constituants fondamentaux des objets et le poids est vu comme une propriété fondamentale de la matière. Nous concluons que les concepts de poids et de densité se différencient chez les enfants au cours du développement et que les concepts des enfants peuvent être présentés dans le contexte de structures de type théories.