

## Transitivity Judgments, Memory for Premises, and Models of Children's Reasoning

MICHAEL CHAPMAN†

*University of British Columbia*

AND

ULMAN LINDENBERGER

*Max Planck Institute for Human Development and Education*

A distributional model of the relation between judgments on transitivity tasks and memory for premise comparisons is proposed, according to which a total population of children solving a transitivity task can be divided into two subpopulations: (a) The *operational* subpopulation consists of all children who infer their transitivity judgments (e.g., stick *A* is longer than stick *C*) from a composition of premise relations (*A* is longer than *B* and *B* is longer than *C*); (b) the *nonoperational* subpopulation consists of all children who infer their transitivity judgments in some other way. In the operational subpopulation, memory for premises should be accurate (because operational composition of premise comparisons depends on the retention of those premises), and transitivity judgments should be correct (because operational composition leads to a correct judgment "by necessity"). In the nonoperational subpopulation, memory for premises should be stochastically independent of transitivity judgments. The assumptions of this distributional model are tested against data on transitive reasoning in 120 first, second, and third graders and found to be reasonable. From the distributional model, an equation is derived allowing the researcher to compute the minimum proportion of operational reasoners required to reject the null hypothesis of independence between judgments and memory in a sample drawn from a mixed (nonoperational + operational) population. Reports of reasoning-remembering independence in previous studies are reinterpreted in light of the present findings. © 1992 Academic Press, Inc.

Ever since the Greeks defined humankind as "the rational animal," the capacity for reasoning has been esteemed as a distinctive feature of human intelligence. In its broadest sense, "reasoning" refers to the explicit consideration of the *reasons* that can be adduced for believing, doing, or valuing one thing over another (Rescher, 1988). As such, it serves the related functions of *decision making* and of *justifying* the decisions which

This article was supported in part by grant number OGP0037334 from the Natural Sciences and Engineering Research Council of Canada. The author would like to thank Darrin Lehman for his comments on an earlier version of the manuscript and Charles Brainerd for his many helpful comments and criticisms. Correspondence concerning this article should be addressed to Ulman Lindenberger, Max Planck Institute for Human Development and Education, Lentzeallee 94, 1000 Berlin 33, Germany.

† Deceased.

are made, whether these functions are performed verbally and collectively or mentally, in the minds of individuals.

In contrast, the psychological investigation of human reasoning generally has been limited to a subset of such functions. For example, Wason and Johnson-Laird (1972) defined the subject of their book as follows: "Our concern is simply this: given a set of assertions, to what extent can the individual appreciate all that follows from them by virtue of logic alone, and remain unswayed by plausible, but fallacious conclusions?" (pp. 1-2). Three implications of this definition are worth noting: (a) The focus is on the *intrapsychic* cognitive processes of individuals; (b) the processes of interest are typically *deductive*—those which follow from a given set of premises to the conclusion(s) which follow from those premises; (c) the question of interest is whether and to what extent the deductive reasoning of real human beings is governed by *logically valid* principles, as opposed to extralogical factors.

Answering this question has been problematic (see Cohen, 1981, and the accompanying commentaries). A major difficulty is the fact that human subjects do not necessarily have veridical insight into their own cognitive processes (Nisbett & Wilson, 1977). This difficulty has been especially acute in developmental studies of human reasoning. Presumably, children are less skilled both in introspection and in the verbal expression of their thoughts than are adults. For these reasons, investigators have attempted to infer the character of children's cognitive processes through a variety of indirect, nonverbal methods (Braine, 1959; Brainerd, 1973; Siegel, 1978). However, the use of such methods has been questioned on the grounds that they might not measure what they were intended to measure originally (Larsen, 1977; Miller, 1976).

This article is devoted to a consideration of one such method for investigating the cognitive processes used by children in common reasoning tasks presumed to require deductive inference. The method is based on an inherent feature of deductive reasoning: In order to make valid inferences from a set of premises, reasoners must at a minimum keep those premises in mind. Further, this *theoretical* dependence of deductive reasoning on an awareness of the required premises should translate in practice into a *statistical* dependence between the conclusion that is presumably derived from a given set of premises and memory for the premises in question; the conclusion should be drawn with greater frequency, if memory for the required premises can be demonstrated (Brainerd & Kingma, 1984, 1985). First, the relevant research on reasoning and memory is reviewed, and the conditions favoring reasoning-remembering dependencies are specified. Next, a distributional model is described which enables the researcher to specify the conditions under which the theoretical dependence between reasoning and memory for premises should result in a

statistical dependence between the stated conclusions and observed memory for premises. Finally, the implications of the model for the interpretation of previous studies as well as for theories of reasoning more generally are discussed.

### REASONING AND MEMORY

Starting with Piaget's (1923/1955, 1924/1928) earliest works, most research on the development of reasoning has been focused on children's increasing grasp of valid forms of inference. Only later did researchers begin to study the retention of relevant background knowledge as a possible limitation on children's reasoning (Bryant & Trabasso, 1971; Roodin & Gruen, 1970). The argument was that young children might fail to draw transitive inferences, not because they failed to grasp the forms of reasoning involved, but because they tended to forget the relevant background information before they could draw the appropriate conclusion. In order to test that hypothesis, Bryant and Trabasso (1971) trained young children to criterion on memory for the premises of a transitive inference before giving them the transitivity task and found that children as young as 4 years of age could solve transitivity problems that typically are not solved until 7 to 8 years of age with Piagetian procedures. This finding appeared to support the conclusion that young children failed the standard procedures because of a memory deficit, not because of any lack of logical reasoning abilities.

That conclusion was controversial from the start. Early criticism centered around the possibility that the transitivity task used by Bryant and Trabasso could be solved by nontransitive forms of reasoning, including perceptual generalization (Youniss & Furth, 1973), "sequential-contiguous" seriation (Breslow, 1981), and labelling (De Broysson-Bardies & O'Regan, 1973; Kallio, 1982; Perner & Aebi, 1985). In reply, Trabasso (1975, 1977) argued that the important question was how children solve transitivity tasks in any case, whether or not their thinking is labelled "true" transitive reasoning. In his view, the results of a number of studies using the Bryant-Trabasso procedures demonstrated that children solve transitivity problems by representing comparison objects in a linear series with spatial properties such that any two members of the series could be compared directly; even if transitive inference were not required in the test phase of the experiment, the integration of premise comparisons into a linear series with transitive properties was indicative of some understanding of transitivity (see also Halford & Kelly, 1984). Subsequent research has suggested, however, that such integration of premises occurs only under specific task conditions, when salient information regarding serial order (Kallio, 1982), graduation of relevant dimensions (Perner & Aebi, 1985), or spatial relations (Chapman & Linden-

berger, 1988) is provided. Therefore, Trabasso's model cannot be considered an explanation of transitive reasoning *in general*, but only of successful performance on transitivity tasks under specific task conditions.

Another line of criticism centered around the statistical relation between children's memory for premises and their judgments on transitive reasoning tasks. Both Halford and Galloway (1977) and Russell (1981) found that some children did not draw transitive inferences even though they recalled the necessary premises, suggesting that failure could not be explained by memory deficits alone. A more radical conclusion was drawn by Brainerd and Kingma (1984), who found that memory for premises and transitivity judgments were stochastically independent, both in their own studies of transitivity and in reanalyses of Halford and Galloway's (1977) and Russell's (1981) data. Whereas most Piagetian and information processing models presuppose a dependency between reasoning and memory, Brainerd and Kingma proposed a "fuzzy trace" model of reasoning, according to which children solved transitivity tasks with informationally impoverished memory traces of the form, "Things get bigger toward the right," rather than with the "precise" premise representations ("A is longer than B," and so on) previously assumed. In another study (Brainerd & Kingma, 1985), they extended this method to other reasoning tasks, including class inclusion, conservation, and probability judgment, as well as transitivity. In each case, children's judgments were found to be stochastically independent of their memory for premises. These results were explained in terms of a functional independence between short-term memory and information processing. Whereas memory for premise information involves precise "verbatim" memory traces, human information processing is adapted to the use of imprecise, gist-like "fuzzy traces." Precise premise information is by-passed in reasoning whenever such fuzzy traces are sufficient for inferring correct judgments.

The novelty of Brainerd and Kingma's argument derives in part from its counterintuitive character. An informal survey of developmental researchers found that 71% believed that memory for the adjacent terms in an  $A > B > C$  transitivity problem should be strongly or moderately correlated with transitivity judgments (Brainerd & Kingma, 1984, pp. 312n-313n). However, the argument for independence of transitivity judgments and memory for premises would be less counterintuitive if children were able to infer correct judgments without the use of transitive reasoning based on adjacent comparisons. No relation between correct judgments and memory for a given set of premises should be found if children can and do infer a correct judgment from an entirely different set of premises. In a comparison between two commonly used versions of the

transitive reasoning task, for example, Chapman and Lindenberger (1988) found that one version (the "alternate" version) could be solved through the use of spatial position cues, because the relevant physical dimension was correlated with spatial position (e.g., comparison objects were arranged from right to left in order of increasing length). The other version (the "standard" version) could not be solved in the same way, because the relevant dimension was uncorrelated with position. Significantly, the "alternate" version was modelled on the task used by Brainerd and Kingma (1984, 1985) for testing the dependency between memory and reasoning. According to Chapman and Lindenberger (1988), one would not expect any reasoning-remembering dependency for that version in any case, because it can be solved without recourse to the set of premises ("A is longer than B," "B is longer than C") for which memory was tested (see also Chapman, 1989).

In reply, Reyna and Brainerd (1989) argued that, even if the Brainerd and Kingma (1984) results could be explained in that way, Chapman and Lindenberger's explanation did not cover all cases; independence between reasoning and remembering had also been found by Halford and Galloway (1977) and by Russell (1981), using tasks that did *not* resemble the "alternate" task used by Chapman and Lindenberger. Additional counterarguments were presented by Reyna and Brainerd (1990). The relevant claims and counterclaims are summarized in the following sections.

#### CONDITIONS FAVORING REASONING-REMEMBERING DEPENDENCIES

Both Brainerd and Kingma (1984) and Chapman and Lindenberger (1988) agreed that reasoning-remembering dependencies should be found under *some* conditions. According to Brainerd and Kingma, such dependencies should be found when retention of the explicit background facts is a "logical precondition for correct reasoning" (1984, p. 364).<sup>1</sup> Similarly, Chapman and Lindenberger argued that reasoning should be related to memory for premises only when subjects actually use a form of reasoning that utilizes the particular premises in question. Therefore, the relevant question is not so much *whether or not* reasoning is dependent on memory, but *under what conditions* subjects are likely to use the form of

<sup>1</sup> Recently, Brainerd and Reyna (1990) have provided a more explicit description of the conditions under which reasoning-remembering dependencies should occur according to fuzzy trace theory. Briefly, such dependencies should be found when memory probes can be answered through reconstructive processing based on the same "fuzzy traces" used in solving reasoning problems. This prediction is discussed in a later section of this paper. For present purposes, the question of interest is what happens when the conditions described by Brainerd and Reyna are *not* met.

reasoning in question. A dependence between the premises and conclusion of that form of reasoning would be expected only under those conditions.

With respect to transitive reasoning, one would expect children's judgments regarding nonadjacent length or weight comparisons (e.g.,  $A > C$ ) to be dependent on memory for adjacent "premise" comparisons ( $A > B$  and  $B > C$ ) only if the children are, in fact, using a form of reasoning in which the conclusion  $A > C$  is derived from the stated premises. If, instead, children infer length or weight from right-left spatial orientation or from some other dimension that happens to be correlated with length or weight, then no dependency between adjacent and nonadjacent comparisons would be expected. At most, a dependency between judgments and knowledge of the respective correlation between dimensions might be expected.

The main significance of the Chapman and Lindenberger (1988) study for present purposes involves the demonstration that the form of reasoning actually used by children in transitivity tasks depends on specific task conditions. *Functional* reasoning—inferring length or weight *as a function* of some other dimension, in this case right-left spatial orientation—should be possible only if length or weight is in fact correlated with spatial orientation (e.g., if the comparison objects are lined up from right to left in front of the child in order of increasing/decreasing length or weight). Under such conditions, children would be expected to infer the nonadjacent comparisons from the spatial correlations rather than from the adjacent comparisons, because the former form of reasoning involves significantly less mental effort than the latter (Chapman, 1987a). If such extraneous correlations do not exist then a consideration of individual premise comparisons should be necessary. Following this line of reasoning, dependence between adjacent and nonadjacent comparisons would be expected only when the extraneous correlations necessary for functional reasoning have been eliminated.

Chapman and Lindenberger (1988) cited three sources of evidence for their conclusion that children used different forms of reasoning depending on whether the relevant dimension in a transitivity task was objectively correlated with right-left spatial orientation (their "alternate" condition) or uncorrelated with spatial orientation (their "standard" condition): (a) Children justified correct answers differently under the two conditions, giving *functional* justifications (e.g., "object  $A$  is longer than object  $C$ , because  $A$  is on the right") in the "alternate" condition and *operational* justifications (" $A$  is longer than  $C$ , because  $A$  is longer than  $B$  and  $B$  is longer than  $C$ ") in the "standard" condition; (b) the task was considerably easier and solved by children at younger ages in the "alternate" condition than in the "standard" control condition; (c) the number of

premise comparisons affected performance only in the “standard,” and not in the “alternate” condition. Chapman and Lindemberger argued that these findings provided converging evidence for the hypothesis that children inferred their judgments as a function of spatial orientation in the “alternate” condition and through a composition of premise relations in the “standard” condition—in other words, that they tended to use the same forms of reasoning in inferring their judgments as in constructing their justifications across conditions. In this view, the number of premise comparisons affected performance in the “standard” but not in the “alternate” condition, because children had to represent the premise comparisons separately in the first, but not in the second.

Those conclusions were challenged by Reyna and Brainerd (1990), who argued that children indisputably can be “in full command of transitivity” without being able to articulate their understanding (cf. Brainerd, 1973, 1977). For this reason, verbal explanations are highly fallible measures of “logical competence,” and the claim that justifications are interesting in their own right (Chapman, 1991) has no bearing on the issue. Besides imposing “an additional performance burden” verbal justifications have no particular advantage over patterns of judgments in assessing the qualitative characteristics of children’s thinking (Reyna & Brainerd, 1990).

Similarly, Reyna and Brainerd questioned whether differential effects of increasing memory load have any relevance for the assessment of competence in reasoning. Memory limitations could constrain reasoning performance in some tasks but not others, depending on whether subjects are able “to access an external store that harbors verbatim information about individual items.” In their view, children’s use of an external array for solving transitivity problems does not imply that they are not competent in transitivity any more than the use of a scratch pad for solving addition problems implies that the persons in question cannot add. In either case, increasing the memory load of problems to be solved will have less of an effect when the external aid is provided than when it is not, but such a result does not have any particular implications for determining what competencies are involved (Reyna & Brainerd, 1990).

### PRO AND CONTRA

We believe that Reyna and Brainerd’s (1990) arguments are flawed. To begin with the issue of children’s verbal justifications as assessment criteria, we do not wish to claim that such justifications make good measures of “logical competence” after all. Instead, we want to distinguish two different kinds of competencies: competence in *intrapsychic inference* by the individual and competence in the *justification* of one’s statements to an interlocutor. Once those competencies are distinguished, one

can ask how they are related. If, as Piaget (1924/1928) once believed, intrapsychic inference develops *from* interpersonal argumentation and justification, then good theoretical reasons exist for expecting a close empirical relation between the two. In order to test this proposition, intrapsychic inference must be assessed independently from verbal justification. A major purpose of the present article is to examine one possible method for doing so: testing the relation between memory for premises and correctness of judgments on a transitive reasoning task.

Similarly, we cannot agree with Reyna and Brainerd that differential effects of increasing numbers of premises are irrelevant for identifying competence in reasoning. If different forms of reasoning are predicted to occur under different task conditions and to be differentially affected by increasing numbers of premises, and if such a pattern of results is actually observed, then those results provide evidence for the hypothesis in question according to any version of the logic of hypothesis testing. In Popperian terms, the hypothesis has survived a test of falsification. Thus, Chapman and Lindenberger's (1988) finding that increasing numbers of premises affected reasoning in their "standard" condition but not in their "alternate" condition provides evidence for the hypothesis by which those results were predicted (that premise comparisons were represented separately in the one condition but not the other). Against Reyna and Brainerd's analogy between children's use of an external array in the transitivity task and the use of a scratch pad for adding a series of numbers, we would argue that the external array condition is more like providing people with sets of objects corresponding to the numbers to be added so that they can obtain the correct sum without adding the numbers at all, but by counting all the individual objects instead. As argued in the case of transitivity tasks featuring an external array, the external aid lessens the memory load of the task by providing subjects with a qualitatively different procedure for obtaining a correct solution.

As this example suggests, Reyna and Brainerd's line of argument rests on the claim that children use a genuine form of transitive inference when provided with an external array as in the Brainerd and Kingma (1984, 1985) studies or in Chapman and Lindenberger's (1988) "alternate" condition. If such were not the case, then Brainerd and Kingma's findings of reasoning-remembering independence would have no bearing on transitive reasoning as such. Accordingly, Reyna and Brainerd (1990) offered three arguments to the effect that children's reasoning is genuinely "logical," even when they are provided with an external array.

The *first* argument was based on a norm-referenced approach: Although children's absolute levels of performance increase with age, the *pattern* of results (reasoning-remembering independence and lack of memory constraints) found by Brainerd and Kingma (1984, 1985) did not



differ by age; thus, the evidence suggests that young children solve transitivity problems in qualitatively the same way as adults. Because adults are generally considered to be competent in transitive reasoning, young children also must be competent in such reasoning, and their lower levels of performance are likely to result from performance deficits rather than from any lack of competence. However, as Reyna and Brainerd admitted a few paragraphs further on, this argument is not decisive because adults themselves do not always think "logically." The question is whether either children *or* adults use genuine transitive reasoning when provided with an external array—or whether that array allows them to use a simpler and qualitatively different form of reasoning. In order to answer this question, some independent criteria for "genuine" transitive reasoning are necessary.

Accordingly, Reyna and Brainerd advanced a *second* line of argument in which Brainerd and Kingma's findings were viewed in relation to the Piagetian concrete-operational criterion of transitive reasoning as the reversible coordination of ordinal relations around a common term. In brief, Reyna and Brainerd argued that (a) two preoperational strategies previously proposed as explanations for young children's high performance on some versions of the transitivity problem—categorical labelling and sequential-contiguous ordering (Breslow, 1981)—cannot account for Brainerd and Kingma's (1984, 1985) results, and therefore (b) the reasoning of children observed in those studies must satisfy the Piagetian reversibility criterion. Curiously, Reyna and Brainerd did not consider the preoperational "functional reasoning" proposed by Chapman and Lindemberger (1988) as a model of the form of inference that children bring to bear on the transitivity problem when provided with an external array. As previously described, this model can account for Brainerd and Kingma's (1984, 1985) results without assuming that children have any representation of an ordered scale based on pairwise relations at all. When children are shown the adjacent terms in an ordered external array, they simply note that the longer of the two is always on the right (or left, as the case may be). Armed with this empirical generalization, they can infer the relation of length between any two terms in the array *as a function of* the right-left spatial relation between the same two terms. They never have to consider more than two terms at a time, and that is why such reasoning does not require any comprehension of an ordered series based on the coordination of pairwise comparisons around a common term. Because Reyna and Brainerd have not ruled out this preoperational model of performance in the external array condition, they cannot claim by default that children's reasoning satisfies the Piagetian criterion for transitivity under such conditions.

The *third* argument of Reyna and Brainerd was based on a procedural

criterion: If children possess a cognitive representation capable of generating correct answers when they are presented with the premises of a transitive inference, then the system can be judged to be "logical" (i.e., a genuine form of transitive reasoning). But that argument fares no better than the first two. For if the premises presented to subjects can be represented by them in different ways, and if the objective situation is such that both kinds of representations can result in correct judgments, then the coincidence of the particular premises as input and correct judgments as output provides no evidence for any particular intervening form of reasoning.

In short, none of Reyna and Brainerd's arguments succeeds in demonstrating that children provided with an external array infer their answers through transitive reasoning understood as the coordination of pairwise comparisons around a common term. However, Trabasso's (1975, 1977) argument is still available to them: Perhaps the important issue is not whether children or adults solve transitivity problems through some idealized criterion for "true" transitive reasoning, but how they actually solve the problems in any case. That position would allow Reyna and Brainerd to maintain the claim that fuzzy trace theory is a general theory of reasoning without saddling themselves with the heavy burden of demonstrating that the forms of reasoning specified in that theory are genuine transitive inferences as commonly understood. Against this claim of generality stands Chapman and Lindenberger's (1988) finding that children really do make transitive inferences based on individual premise comparisons under some conditions—indeed, that the form of reasoning described in fuzzy trace theory occurs only under very special conditions. But as Reyna and Brainerd (1990) pointed out, Chapman and Lindenberger's conclusions are based on the disputed assumption that children in their "standard" condition inferred their judgments in a manner roughly consistent with their verbal justifications. Perhaps, instead, they inferred their judgments as described in fuzzy trace theory and justified them only after the fact with reference to individual premise comparisons.

The major goal of this study was to provide additional tests of the hypothesis that children infer transitive judgments through a composition of individual comparisons when the relevant physical dimension is uncorrelated with spatial orientation. The method used is that pioneered by Brainerd and Kingma (1984, 1985), who found that transitivity judgments and memory for premises were stochastically independent when children were tested under the external array condition. The question addressed in this study is whether such independence between reasoning and remembering is found when children are *not* provided with such an external array—that is, under Chapman and Lindenberger's "standard" condition.

Another important question is under what conditions reasoning-remembering dependencies, if they do exist, would be detected by the researcher. Brainerd and Kingma (1984, 1985) give two major reasons why the absence of such a dependency in their analyses is unlikely to have resulted from a lack of statistical power: (a) The data entering into each statistical test were based upon a large number of protocols, so their null findings should not have resulted from false negative measurement error. (b) The fact that a reasoning-remembering dependence was found for *some* tasks (including numerical invariance, mental arithmetic, and conservation) in which precise premise information presumably was required for correct reasoning, argues against the possibility that the analyses were incapable of detecting reasoning-remembering dependencies in principle. Evaluation of those claims necessitates some consideration of the *form* that reasoning-remembering dependencies is likely to take when it does occur. In the next section, this problem is addressed in the context of Piaget's theory of concrete operations.<sup>2</sup>

#### THEORETICAL ASSUMPTIONS AND A DISTRIBUTIONAL MODEL

According to Piaget (1947/1950, 1949/1972a), concrete operational reasoning is "operational" in two senses: (a) the task information is assimilated by the subject through first-order "operations" originating from interiorized actions; and (b) inferences beyond the information given are drawn by means of a second-order operation which acts by "composing" two or more of those first-order operations. In the transitivity of length task, the first-order operations are the interiorized actions involved in comparing the lengths of *A* and *B* and those of *B* and *C*, and the second-order operation is the *composition* of those two comparisons. From the composition of the *AB* and *BC* comparisons, the *AC* comparison can be

<sup>2</sup> Our references to Piaget's theory of concrete operations do not imply that we accept the traditional interpretation of his structural stage theory, according to which children are expected to reason consistently across tasks and domains within a given developmental stage. The argument is twofold: (a) Piaget himself did not believe in such consistency and the traditional interpretation as stated above arose in the assimilation of Piaget's theory into mainstream developmental psychology (Chapman, 1987a, 1988), and more importantly, (b) one must acknowledge the context dependency of children's forms of reasoning, *quite apart* from what Piaget may or may not have believed. The fact that children are capable of a higher-level form of reasoning does not mean they will use it under all circumstances whatever. For reasons of cognitive economy, one might expect them to continue using the lower-level form of reasoning in situations in which it is sufficient for the task at hand. For example, inferring weight from size is quite sufficient in a situation in which the objects involved are in fact of equal density. The developmental problem for children is to recognize when a lower-level form of reasoning is *not* appropriate and to develop a higher-level form which is.

generated. So defined, operational reasoning can be used for either purposes of deductive inference or justification, depending on whether the primary movement is from the premises to the conclusion (“ $A > B$  and  $B > C$ , so  $A > C$ ”) or from the conclusion to the premises which justify it (“ $A > C$ , because  $A > B$  and  $B > C$ ”), respectively. One might expect operational competence in justification to be related in some degree to operational competence in intrapsychic inference because both have either the same form or, according to some theories of *sociogenesis*, the second might develop from the first (Chapman, 1991, in press).

Another relevant characteristic of operational reasoning is the “logical necessity” associated with it. Piaget frequently wrote that concrete operational thinking was characterized by a “sense of necessity” (e.g., Piaget, 1971). Evidence that both children and adults do associate concrete-operational reasoning with some form of logical necessity has been reported by Miller (1986). However, one would be mistaken to equate the sense of necessity characteristic of concrete operations with “formal” or “analytic” necessity—i.e., with the requirement that the meaning of a conclusion is already contained in the premises. The fact that concrete operational reasoning involves some empirical content has often been noted (e.g., Brainerd, 1973; Hall & Kaye, 1978), but this does not imply that such reasoning does not also involve necessity, only that necessity is less than analytic in character.

Piaget’s solution to this problem was to ascribe a weaker sense of necessity to concrete operations than the formal, analytic necessity characteristic of formal operations. In his posthumously published book on the development of necessity (Piaget, 1983/1987a), he pointed out that the sense of necessity which develops at the stage of concrete operations has a modal character: The truth of an assertion is *necessary*, if its negation is *impossible* (see also Piaget, 1977/1986). Because the range of possibilities considered by concrete operational thinkers is primarily empirical (Piaget, 1981/1987b), there is a corresponding sense of necessity that depends upon knowledge of what is empirically possible and impossible. Anyone with an empirical knowledge of length as a physical dimension will recognize that  $C$  *cannot* be longer than  $A$ , if  $A$  is longer than  $B$  and  $B$  is longer than  $C$ . Therefore,  $A$  *must* be longer than  $C$ .

The origin of concrete operational necessity is a question to which we return later. For present purposes, the important point is the following: The element of necessity in concrete operational reasoning guarantees that certain “premise” information will uniquely determine the resulting judgment. In terms of transitive reasoning, these considerations imply that a reasoner who is provided with accurate information that  $A$  is longer than  $B$  and  $B$  is longer than  $C$  and who, in fact, uses an operational form of inference will always return the judgment that  $A$  is longer than  $C$ . In

theory, accurate premise information together with operational inference should always lead to correct judgments; in practice, some instances of misunderstanding, faulty inference or misspeaking, are likely to occur with some small probability. A similar argument was made by Scribner (1977) for syllogistic reasoning: If subjects infer their judgments from a deductive composition of premises, then those judgments should invariably be correct, except for a small proportion of errors of comprehension and production just mentioned. Consistent with this hypothesis, she found that, across seven studies of syllogistic reasoning conducted in Mexico and Africa, the proportion of theoretical reasons associated with wrong answers averaged about 1%. Of 600 problems in the African study, not a single case was found in which a subject giving a deductive justification had a wrong answer. A low incidence of children giving incorrect judgments and deductive justifications in a transitivity task was reported by Russell (1981).

To summarize the argument for transitivity: if "operational reasoning" is defined as reasoning based on a composition of premise relations, then children inferring their transitivity judgments operationally should (a) have accurate memory for premises (otherwise, they would not be able to make operational inferences) and (b) exhibit correct judgments (because operational inference, given accurate premise information, "necessarily" leads to a correct judgment). More generally, a population of children solving a transitivity task is potentially divisible into "nonoperational" and "operational" subpopulations, based on their respective ways of inferring judgments on that task. (Note that the two subpopulations are defined in a *task-specific* manner; we do not assume that children solving one task operationally will necessarily solve other tasks in the same way. As used here, the term "operational" describes a form of reasoning used by a child in a particular situation, not a general trait-like "competence" attributed to the child as an individual.) The distributional assumptions underlying the nonoperational and operational subpopulations, so understood, are illustrated in Table 1.

In the *nonoperational* subpopulation, the underlying assumption is that memory for premises and transitivity judgments are stochastically independent. This assumption follows from the foregoing argument that memory and judgments should be related only when subjects are reasoning operationally. Mathematically, this stochastic independence implies that the probabilities of the joint events determined by the pairing of accurate or inaccurate memory with correct or incorrect judgments (i.e., the probabilities  $p_{mj}$ ,  $p_{\bar{m}j}$ ,  $p_{m\bar{j}}$ ,  $p_{\bar{m}\bar{j}}$ ) should be equal to the products of the corresponding marginal probabilities ( $p_{mj} = p_m p_j$ , etc.). In the *operational* subpopulation, memory for premises and judgments are assumed to be dependent, because memory should always be accurate and judgments

TABLE I  
DISTRIBUTIONAL MODELS OF NONOPERATIONAL AND OPERATIONAL SUBPOPULATIONS

Memory for premises	Judgments		Total
	Correct	Incorrect	
Nonoperational subpopulation			
Accurate	$(p_{m \bar{o}} p_{j \bar{o}})/p_{\bar{o}}$	$(p_{m \bar{o}} p_{\bar{j} \bar{o}})/p_{\bar{o}}$	$p_{m \bar{o}}$
Inaccurate	$(p_{\bar{m} \bar{o}} p_{j \bar{o}})/p_{\bar{o}}$	$(p_{\bar{m} \bar{o}} p_{\bar{j} \bar{o}})/p_{\bar{o}}$	$p_{\bar{m} \bar{o}}$
Total	$p_{j \bar{o}}$	$p_{\bar{j} \bar{o}}$	$p_{\bar{o}}$
Operational subpopulation			
Accurate	$p_o$	0	$p_o$
Inaccurate	0	0	0
Total	$p_o$	0	$p_o$

*Note.* Cell entries represent the marginal probabilities for the events indicated, under the distributional assumptions described in the text. All probabilities are expressed relative to the total population.

should always be correct. The relation between the two variables is thus deterministic, rather than stochastic. To be sure, some cases of inaccurate memory and incorrect judgments are likely to be observed in any sample drawn from this population because of measurement, comprehension, or production errors. But in the ideal, theoretical model depicted in Table 1, these sources of error are overlooked.

From the earliest age at which children begin to develop operational reasoning with respect to a given task, any random sample of children tested on that task will be drawn from a total population composed of nonoperational and operational subpopulations in proportions ranging between 0 and 1.00. Further, a stochastic dependence between memory and judgments is expected to occur in the total population, resulting from the admixture of an operational subpopulation (in which judgments and memory are related deterministically) to a nonoperational subpopulation (in which judgments and memory are stochastically independent). If  $p_o$  is the proportion of children using operational reasoning in the total population, then the distributional model of that total population can be determined as shown in Table 2.

This distributional model makes it possible to estimate the minimum value of  $p_o$  necessary to reject reasoning-remembering independence for various values of  $p_m$  and  $p_j$ , thereby providing a method for checking previous reports of independence between reasoning and remembering in children. Before applying this method, however, it is prudent to test the "reasonableness" of the model with reference to empirical findings.

TABLE 2  
 DISTRIBUTIONAL MODEL OF A TOTAL POPULATION FORMED BY A MIXTURE OF  
 NONOPERATIONAL AND OPERATIONAL SUBPOPULATIONS

Memory for premises	Judgments		Total
	Correct	Incorrect	
Accurate	$(p_{m \bar{o}} p_{j \bar{o}})/p_{\bar{o}} + p_o$	$(p_{m \bar{o}} p_{j \bar{o}})/p_{\bar{o}} + 0$	$p_m$
Inaccurate	$(p_{\bar{m} \bar{o}} p_{j \bar{o}})/p_{\bar{o}} + 0$	$(p_{\bar{m} \bar{o}} p_{j \bar{o}})p_{\bar{o}} + 0$	$p_{\bar{m}}$
Total	$p_j$	$p_{\bar{j}}$	1.00

*Note.* Cell entries represent the joint probabilities and marginal probabilities which result from adding the respective probabilities in the nonoperational and operational subpopulations given in Table 2.

### TESTING THE MODEL

In order to test the assumptions underlying the distributional model described in the preceding section, the data reported by Chapman and Lindenberger (1988) will be reanalyzed for reasoning-remembering dependencies. For this purpose, only the data from their "standard" transitivity tasks will be used. This was the version of the transitivity task in which length and weight were *not* correlated with spatial position. It was therefore the only version in which reasoning-remembering dependencies might be expected to occur. First, Chapman and Lindenberger's total sample will be analyzed for reasoning-remembering dependence after the manner of Brainerd and Kingma (1984, 1985). Then, the predictions generated by the present model for operational and nonoperational subpopulations considered separately will be tested.

As described in the original article, Chapman and Lindenberger gave 3-, 4-, and 5-term transitivity-of-length tasks and 3- and 4-term transitivity-of-weight tasks to 120 first, second, and third graders with mean ages of 7.06, 7.91, and 8.95, respectively. The "standard" transitivity-of-length tasks were administered with color-coded sticks of varying lengths. First, adjacent "premise" comparisons were presented by placing the two sticks upright next to each other on the table so that the difference between lengths was clearly visible. When all premise comparisons had been presented in this way, memory for premises was checked by reintroducing each pair of sticks, this time with their ends hidden from view so that differences in length could not be seen, and asking children if they could remember which stick was longer. If they could not remember, or if their memory was inaccurate, the length differences were demonstrated once again and the memory check repeated. The second time, however, failures were not corrected. Instead, the experimenter proceeded to pose the transitivity problem, showing children the end terms of the  $n$ -term

length series and asking children which of them was longer. Again, sticks were presented with their ends hidden so that length differences could not be seen. Finally, children were asked to explain their answers. Weight tasks were administered in an analogous manner, except that comparison objects were color-coded balls differing in size and weight, and weight differences were demonstrated by the experimenter with a balance scale. A complete description of the procedures is provided in Chapman and Lindenberger (1988).<sup>3</sup>

<sup>3</sup> Because our argument relies in part on the results of Chapman and Lindenberger (1988), it is important to consider some recent criticisms of that study by Howe and Rabinowitz (1990):

a. Chapman and Lindenberger were said to provide no direct evidence for a dependence between reasoning and memory for premises, because "no memory measures were taken."

b. It is not clear how children could have given an operational explanation on Chapman and Lindenberger's "alternate" task, because the comparison objects were not distinguished by color as they were on the "standard" task.

c. Chapman and Lindenberger's finding that the difficulty of reasoning varied with the number of comparison objects for the "standard" task, but not for the "alternate" task was said to be an artifact of their design, because children in the "standard" tasks were tested on separate 3-, 4-, and 5-term problems whereas children in the "alternate" tasks were questioned about 3-, 4-, and 5-term (e.g., *AC*, *AD*, and *AE*) comparisons on a single 5-term display. In addition, children's scores on the standard tasks were based on only one response, whereas scores on the "alternate" tasks were sometimes based on multiple responses (i.e., 3-term scores were based on *AC*, *BD*, and *CE* comparisons, and 4-term scores were based on *AD* and *BE* comparisons).

Although some of Howe and Rabinowitz's remarks are well taken, others are based on misconceptions of the methods and goals of the Chapman and Lindenberger study. In any event, we believe that the essential conclusions of that study are not negated by the factors mentioned.

a. The claim that Chapman and Lindenberger (1988) included no measures of children's memory for premises is not true. The memory probes are described in their methods section, and the data bearing on the relation between judgments and memory for premises are described in this article.

b. It was quite simple for children to give operational explanations on the "alternate" task, for example, by stating that *A* is longer than *C*, "because this one (pointing to *A*) is longer than that one (pointing to *B*), and that one (*B*) is longer than that one (*C*)." As noted by Chapman and Lindenberger (1988, p. 545), some children indeed gave operational justifications as well as "functional" justifications, but in each case the functional explanation was preferred. If anything, it should have been easier for children to give operational explanations on the alternate task, because all the comparison objects were present and they did not have to remember the colors of intermediate terms which were *not* perceptually present, as they had to do in the standard task. Although it is certainly possible that different colors in Chapman and Lindenberger's alternate task might affect children's answers in some way, it is by no means obvious a priori that such a control would alter their major results.

c. Howe and Rabinowitz provide no evidence for their assertion that presenting 3-, 4-, and 5-term "alternate" tasks separately would eliminate the difference in difficulty patterns for "standard" versus "alternate" tasks in the Chapman and Lindenberger study. In fact, the necessary control already has been done by Brainerd and Kingma (1984), who presented



In the original study, children's answers were classified according to a criterion involving both judgments and verbal justifications. However, judgments and justifications were also coded separately, making possible an analysis of the relation between memory and judgments independent of justifications in a manner comparable to that of Brainerd and Kingma (1984, 1985). The cross-classification of children's memory for premises and transitivity judgments for each type of transitivity problem is presented in Table 3. In that analysis, children's memory for premises was considered "accurate" if they successfully recalled all the premise comparisons necessary for operational reasoning (i.e., all the adjacent comparisons). The  $\chi^2$  test of independence for each type of problem is also provided.

As indicated in the table, memory for premises was independent of judgments for four of the five transitivity tasks. Only for the 3-term transitivity of length task was memory significantly associated with judgments. Thus, the overall result would appear to be consistent with Brainerd and Kingma's (1984) finding of independence between children's memory and transitivity judgments.<sup>4</sup> However, the fact that an associa-

---

3- and 5-term tasks separately and obtained the same pattern of results (no difference in difficulty) as reported by Chapman and Lindemberger for the "alternate" task. The results of the "alternate" task are not at issue. The point of the Chapman and Lindemberger study was the demonstration that at least one version of the transitivity task exists (i.e., the "standard" version) in which the number of terms *does* make a difference and for which (as shown in this article) reason and remembering are related. As for Chapman and Lindemberger's scoring of 3- and 4-term "alternate" problems with multiple responses, it was, in fact, a conservative procedure. Children were given a "pass" on these problems only if they passed *all* the comparisons in question (e.g., the *AD* and *BD* comparisons for the 5-term problem). If anything, this procedure would tend to underestimate children's performance, because the probability of passing them all must be less than (or equal to) the probability of passing each one individually. Therefore, any effect of using multiple responses would be to minimize Chapman and Lindemberger's reported finding that the "alternate" tasks were passed at a much higher rate than the "standard" tasks.

<sup>4</sup> Although Brainerd and Kingma (1984, 1985) used likelihood ratios in testing the independence of judgments and memory, such tests become equivalent to the Pearson  $\chi^2$  tests used in this paper for large  $N$ . Note that their published description of the likelihood ratio test contained some errors (C. J. Brainerd, personal communication, June 16, 1989). Briefly, the null hypothesis of equality between (a) the conditional probability of a correct judgment, given accurate memory, and (b) the simple probability of a correct judgment (i.e.,  $p_{jm} = p_j$ ) was said to imply that the joint probability of a correct judgment and accurate memory was equal to the joint probability of a correct judgment and inaccurate memory ( $p_{mj} = p_{\bar{m}j}$ ). The likelihood ratio test was derived from the latter equation (Brainerd & Kingma, 1985, p. 223). In fact,  $p_{mj} = p_{\bar{m}j}$  does *not* follow from  $p_{jm} = p_j$ , as can be verified from the formula for conditional probabilities. Further, it is unnecessary to test the hypothesis  $p_{jm} = p_j$  separately from the hypothesis that  $p_{jm} = p_m$  as stated by Brainerd and Kingma, because  $2 \times 2$  tables have only 1 degree of freedom and therefore each of these equalities implies the other.

TABLE 3  
CROSS CLASSIFICATION OF CHILDREN'S MEMORY FOR PREMISES AND  
TRANSITIVITY JUDGMENTS

Memory for premises	Judgments		$\chi^2(1)$
	Correct	Incorrect	
	3-term transitivity of length		
Accurate	80	23	8.00*
Inaccurate	7	10	
	4-term transitivity of length		
Accurate	51	27	0.00
Inaccurate	27	15	
	5-term transitivity of length		
Accurate	31	16	2.13
Inaccurate	37	36	
	3-term transitivity of weight		
Accurate	72	28	0.64
Inaccurate	12	8	
	4-term transitivity of weight		
Accurate	61	33	0.01
Inaccurate	16	10	

Note.  $\chi^2$  values corrected for continuity (Yates' correction).

\*  $p < .005$ .

tion was found even for one of the five tasks argues for caution in drawing any general conclusions before completing the second analysis of operational and nonoperational subpopulations separately.

The major assumptions of the distributional model described in the previous section can be summarized as follows: (a) the total population from which any sample of children is drawn can be partitioned into nonoperational and operational subpopulations for each task; (b) memory for premises and transitivity judgments are stochastically independent in the nonoperational subpopulation; and (c) memory is always accurate and judgments are always correct in the operational subpopulation, ignoring the effects of measurement and processing error. In order to test the reasonableness of these assumptions, some method of estimating membership in the operational subpopulation is required. As described below, the most important of the three assumptions is the third one, leading to a prediction of reasoning-remembering dependencies in the operational subpopulation. Therefore, the method of estimating membership in that subpopulation should be characterized by low false-positive measurement

error: In order to test a theoretical characteristic of this subpopulation, one should have confidence that the sample selected actually belongs to that population. Less important for this purpose is the possibility that some subjects belonging to the desired subpopulation are *not* selected (false-negative measurement error).

Although we agree that children's verbal justifications do not necessarily reflect the forms of their intrapsychic inference, we believe that verbal justifications are appropriate for the specific purpose of estimating membership in the operational subpopulation for the following reasons: As mentioned previously, the use of verbal justifications frequently has been criticized as involving high false-negative measurement error. Brainerd (1973, 1977) and others (e.g., Braine, 1959; Gruen, 1966; Siegel, 1978) have argued that children might fail to give operational justifications, not because they cannot (or do not) reason operationally, but because they cannot put their reasoning into words. However, false-negative error (not sampling all children using operational reasoning) is less serious a problem for present purposes than false-positive error (sampling children who in fact did not use operational reasoning). In this connection, both the critics and defenders of verbal methods agree that using verbal justifications as criteria for concrete operational reasoning is likely to involve low false-positive error (Brainerd, 1973; see also Reese & Schack, 1974).<sup>5</sup> Accordingly, children were judged as belonging to the operational subpopulation if they justified their transitivity judgments with reference to all the premise comparisons between adjacent terms in the series. Such operational justifications were coded with sufficient interrater reliability (95% agreement) in Chapman and Lindemberger's (1988) study to permit their use for this purpose.

The assumption that only accurate memory for premises and correct judgments would be found in the operational subpopulation was tested by

<sup>5</sup> One might object to this conclusion regarding low false-positive measurement error with the following argument: Children could infer a correct answer by one form of reasoning and "rationalize" their answers after the fact through some other form of reasoning, as has been reported for adults (e.g., Nisbett & Wilson, 1977). This type of error would pose a problem of interpretation if the predicted dependence between judgments and memory for premises in the operational subpopulation failed to materialize. Then one would not know if this lack of dependence was attributable to the fact that (a) children inferred their judgments operationally but did so (somehow!) without accurate memory for premises and/or without correct judgments, or (b) children classified as "operational" because of their judgments did not use operational reasoning in drawing their inferences. The problem of false-positive measurement error would be less problematic if the predicted dependence *were* obtained—a result that would be consistent with the conclusions that the (a) false positive error rate was low and (b) predicted dependence between children's judgments and memory for premises *is* found when children infer their judgments operationally. In fact, the second of these two results was obtained, suggesting that false-positive errors were not a problem.

TABLE 4  
CROSS CLASSIFICATION OF CHILDREN'S MEMORY FOR PREMISES AND TRANSITIVITY  
JUDGMENTS, BY FORMS OF JUSTIFICATION

Memory for premises	Judgments by justifications			
	Nonoperational		Operational	
	Correct	Incorrect	Correct	Incorrect
	3-term transitivity of length			
Accurate	25	23	55	<i>0</i>
Inaccurate	7	10	<i>0</i>	<i>0</i>
	4-term transitivity of length			
Accurate	37	27	14	<i>0</i>
Inaccurate	27	15	<i>0</i>	<i>0</i>
	5-term transitivity of length			
Accurate	28	16	3	<i>0</i>
Inaccurate	37	36	<i>0</i>	<i>0</i>
	3-term transitivity of weight			
Accurate	38	28	34	<i>0</i>
Inaccurate	12	8	<i>0</i>	<i>0</i>
	4-term transitivity of weight			
Accurate	52	33	9	<i>0</i>
Inaccurate	16	10	<i>0</i>	<i>0</i>

*Note.* Error cells indicated with italics.

partitioning each memory-by-judgments cross-classification matrix shown in Table 4 into two submatrices according to children's forms of justification. These submatrices are shown in Table 4. One submatrix (labelled "nonoperational") included children who gave nonoperational justifications or no justifications at all. This group was assumed to include children belonging to the nonoperational subpopulation as well as an unknown proportion of children from the operational subpopulation who nevertheless did not give operational justifications (i.e., the false negatives). The other submatrix included only children who gave operational justifications, that is, justifications referring to all the relevant premise comparisons (e.g., "A is longer than C, because A is longer than B and B is longer than C"). This group was assumed to include only children belonging to the operational subpopulation (i.e., a negligible number of false positives).

From the distributional model described previously, the following predictions regarding the two submatrices shown in Table 4 were derived: (a) No significant association between memory and judgments should be

found in the nonoperational submatrix, under the assumption that the false-negative error rate was reasonably low, and (b) all observations in the operational submatrix should fall into the cell defined by accurate memory and correct judgments, assuming a negligible false-positive error rate. The other three cells in the operational submatrix are "error cells," which should be empty according to theory. For several reasons, the second of these predictions is the more meaningful one. Not only does the first prediction involve the dubious practice of accepting the null hypothesis of independence as well as an assumption regarding the low rate of false negatives, but as we shall see presently, meaningful relations between memory and judgments might not always be detected even if they exist. Consistent with the second prediction, all observations in the operational submatrix fell in the cell defined by accurate memory and correct judgments as shown in Table 4. The error cells (indicated by italics in the table) were empty, as predicted.

Test statistics relevant to both predictions are presented in Table 5. The prediction of independence between memory for premises and transitivity judgments in the nonoperational submatrices of each type of transitivity task was tested with  $\chi^2$  tests of association, and the prediction that no observations in the operational submatrix should fall into the error cells defined by inaccurate memory or incorrect judgments was tested with the PRE-statistic, a measure of Proportional Reduction of Error (Hildebrand, Laing, & Rosenthal, 1977; von Eye & Brandtstädter, 1988). The significance of PRE was tested by the normal approximation to the binomial as recommended by von Eye and Brandtstädter (1988). According to this procedure, the null hypothesis that PRE equals 0 (i.e., that the number of

TABLE 5  
TEST STATISTICS FOR THE NONOPERATIONAL AND OPERATIONAL SUBMATRICES

Transitivity task	Submatrix		
	Nonoperational $\chi^2(1)$	Operational <i>PRE</i>	<i>z</i>
Length			
3-term	0.24	1.00	4.89**
4-term	0.21	1.00	2.76**
5-term	1.38	1.00	1.21
Weight			
3-term	0.00	1.00	3.87**
4-term	0.00	1.00	1.92*

Note. Both  $\chi^2$  and *z* values corrected for continuity.

\*  $p < .05$ .

\*\*  $p < .001$ .

observations falling in the error cells is equal to the number expected by chance) is tested against the alternative hypothesis that PRE is greater than 0 (that the number of observed errors is *less* than that expected by chance). In the present case, a separate prediction analysis was performed for each transitivity task shown in Table 4. Expected cell frequencies were determined from the marginal distributions for operational versus nonoperational justifications, correct versus incorrect judgments, and accurate versus inaccurate memory for premises. Thus, the probability of an observation occurring in a given cell  $ijk$  was derived from the respective marginal probabilities corresponding to those dimensions ( $p_{ijk} = p_i p_j p_k$ ). This null distribution is based on the assumption of independence among the three dimensions and forms the background against which predictions of specific dependencies are tested (von Eye & Brandtstädter, 1988).

As indicated in the table, the  $\chi^2$  tests were nonsignificant for all the transitivity tasks, but PRE was significantly greater than 0 at a Bonferroni criterion of .001 (= .005/5) for three of the five tasks. The fact that PRE was equal to 1.00 for each task means that prediction was perfect (no errors were observed) in each case, but the numbers of children in the operational submatrices for the 5-term length task and the 4-term weight task were too small to reach significance. The fact that prediction was perfect in the total sample implies that the observed relation between judgments and memory was obtained within each of the three age groups as well. Therefore, the results described in Tables 4 and 5 are not artifacts of combining data from children of different ages. The numbers of operational reasoners in the first-, second-, and third-grade groups, respectively, were as follows: 3-term length task—10, 21, 24; 4-term length task—1, 5, 8; 5-term length task—0, 1, 2; 3-term weight task—3, 14, 17; 4-term weight—1, 4, 4.

Two general observations can be made about these results. First, they are indeed consistent with the predictions derived from the distributional model summarized in Table 2: (a) For children giving nonoperational justifications, memory was unrelated to judgments; (b) for children giving operational justifications, memory was always accurate and judgments were always correct. Although the latter finding might appear to be intuitively obvious, it was by no means inevitable. If Brainerd and Kingma's (1984, 1985) hypothesis of independence between reasoning and remembering were correct, then one might expect some children to have answered the memory probes inaccurately but to have guessed or reconstructed the relevant premises in giving an operational justification. And if the cognitive processes by which children's judgments are generated are not necessarily reflected in their verbal justifications (Braine, 1959; Brainerd, 1973), then one might expect at least some children giving op-

erational justifications to have incorrect judgments. Although the opposite combination (correct judgment and nonoperational justification) might be considered more likely to occur, the combination of incorrect judgment and operational justification has indeed been reported to occur with low frequency under some task conditions (e.g., Russell, 1981), as mentioned previously.

The second observation to be made is that the reasoning-remembering dependencies indicated by the results shown in Tables 4 and 5 were not detected by the  $\chi^2$  tests for the total sample shown in Table 3, except for the 3-term transitivity of length task. This result raises doubts about the power of tests of association for detecting reasoning-remembering dependencies, even if they do exist. This question of statistical power will now be examined more closely.

#### CONDITIONS FOR REJECTING THE HYPOTHESIS OF INDEPENDENCE

Under the assumption that the distributional model shown in Table 2 characterizes the total population from which children are sampled, the minimum proportion of operational reasoners ( $P_o^*$ ) necessary to reject the null hypothesis of independence in tests of association can be calculated. This is accomplished by substituting the values given in Table 2 into the familiar computational formula for a  $2 \times 2$   $\chi^2$  test of independence, simplifying, and solving for  $P_o$ . This procedure yields the following formula:

$$P_o^* = \left[ \left( \frac{N p_{\bar{j}} p_{\bar{m}}}{\chi^2_{\alpha} p_j p_m} \right)^{1/2} + 1 \right]^{-1} \quad (1)$$

where  $P_o^*$  is the minimum proportion of true operational reasoners in a sample of  $N$  children, at the  $\alpha$  level of significance, for the hypothetical population probabilities  $p_j$  and  $p_m$  (where  $p_{\bar{j}} = 1 - p_j$  and  $p_{\bar{m}} = 1 - p_m$ ). The values of  $P_o^*$  obtained by this method for  $\alpha < .05$ ,  $N = 100$ , and various combinations of  $p_m$  and  $p_j$  between .20 and .90 are presented in Table 6.

It should be emphasized that the values of  $P_o^*$  in Table 6 are *minimum* values obtained under the ideal assumption that the sample conforms perfectly to the theoretical model in Table 2. In other words, the effects of measurement error or other sources of random error are ignored. Given the existence of error in any actual sample, other deviations from the theoretical model, or  $N < 100$ , the critical values are likely to be larger. Nevertheless, one is impressed at how large the values of  $P_o^*$  in Table 6

TABLE 6  
 MINIMUM PROPORTION OF OPERATIONAL REASONERS NEEDED TO REJECT NULL  
 HYPOTHESIS OF INDEPENDENCE BETWEEN MEMORY AND JUDGMENTS

Probability of accurate memory	Probability of correct judgment							
	.20	.30	.40	.50	.60	.70	.80	.90
.20	.05	.07	.08	.09	.11	.14	.17	.23
.30	.07	.08	.10	.12	.14	.17	.21	.28
.40	.08	.10	.12	.14	.17	.20	.25	.35
.50	.09	.12	.14	.17	.20	.24	.29	.38
.60	.11	.14	.17	.20	.23	.27	.33	.42
.70	.14	.17	.20	.24	.27	.32	.38	.48
.80	.17	.21	.25	.29	.33	.38	.44	.55
.90	.23	.28	.35	.38	.42	.48	.55	.64

Note. Table entries calculated from Eq. (1) for  $\chi^2_{.05}(1) = 3.841$  and  $N = 100$ .

are, even under the ideal conditions described. For values of  $p_m$  and  $p_j$  greater than .50, nearly one-fifth of the sample could be true operational reasoners without rejecting the null hypothesis of independence between memory and judgments. For  $p_m$  and  $p_j$  greater than .70, more than one-third of the sample would have to be operational to reject the null hypothesis. In short, tests of association are not particularly sensitive for detecting the type of reasoning-remembering dependency that exists in operational reasoners, even under the overly optimistic assumption of negligible measurement error.

As an illustration of the limitations on statistical power represented by the figures in Table 6, the critical proportions ( $P_o^*$ ) necessary to reject the null hypothesis of independence for the Chapman and Lindenberger (1988) data were calculated from Eq. 1. The resulting values for  $P_o^*$ , the observed proportions  $P_m$  and  $P_j$  used in calculating  $P_o^*$ , as well as the observed proportion  $P_e$  of children giving operational justifications are given in Table 7. For each transitivity task,  $P_o^*$  represents the minimum proportion of true operational reasoners necessary to reject the null hypothesis of independence in  $\chi^2$  tests of association under the distributional assumptions described in preceding sections. For present purposes,  $P_e$  may be regarded as a rather conservative estimate of the proportion of operational reasoners in the population. The main result shown in the table is the fact that  $P_e$  exceeds  $P_o^*$  only for the 3-term transitivity of length task. Therefore, one would expect a significant association between memory and judgments to be found only for this task, precisely the result that was obtained (see Table 3). Despite the fact that the estimated proportion of operational reasoners was quite high for some of the other tasks (at least 28% for the 3-term transitivity of weight task), significant  $\chi^2$  tests would not be expected because  $P_e < P_o^*$ .



TABLE 7  
OBSERVED PROPORTIONS OF ACCURATE MEMORY ( $P_m$ ), CORRECT JUDGMENT ( $P_j$ ), AND OPERATIONAL EXPLANATIONS ( $P_e$ ) WITH CRITICAL PROPORTION OF OPERATIONAL REASONERS ( $P_o$ )

Transitivity task	Observed proportions			Critical Proportion <sup>a</sup> $P_o^*$
	$P_m$	$P_j$	$P_e$	
<b>Length</b>				
3-term	0.86	0.73	0.46	0.42
4-term	0.65	0.65	0.12	0.25
5-term	0.39	0.57	0.03	0.14
<b>Weight</b>				
3-term	0.83	0.70	0.28	0.38
4-term	0.78	0.64	0.08	0.31

<sup>a</sup> Calculated from Eq. 1 for  $\chi^2_{.05}(1) = 3.841$ ,  $N = 120$ , and observed values of  $P_m$  and  $P_j$ .

### REINTERPRETING PREVIOUS RESULTS

The foregoing analyses provide a basis for reinterpreting previous reports of independence between reasoning and remembering, beginning with the studies of transitive reasoning reviewed by Brainerd and Kingma (1984). In their reanalysis of data reported by Halford and Galloway (1977), Russell (1981), and Kingma (1981), Brainerd and Kingma found no statistically reliable dependencies between children's transitivity judgments and their memory for premise comparisons. However, they note that such statistical independence could have resulted from methodological problems in all three of those studies (Brainerd & Kingma, 1984, pp. 328–329). In each case, the procedures used to measure judgments and memory were in some way incomparable, introducing significant sources of measurement error into one or the other of the two variables. In Halford and Galloway's study, a horizontal-vertical visual illusion was employed in presenting transitivity questions but not memory probes. Therefore, transitivity judgments and memory for premises might have been unrelated because judgments, but not memory, would have depended in part on the ability to resist the illusion. In Russell's and in Kingma's studies, memory for premises was inferred from children's verbal explanations of their judgments. Thus, memory for premises as assessed would have depended not only on accurate memory, but also on whatever verbal skills and reasoning abilities are involved in formulating a verbal explanation. Transitivity judgments and memory for premises might have been unrelated in these studies because of the measurement error associated with this use of explanations as a measure of memory.

Because of the methodological problems associated with these previous studies, Brainerd and Kingma (1984) conducted a series of their own

transitivity studies in which memory for premises and transitivity judgments were assessed with comparable methods. Both memory and transitivity judgments were elicited by asking children which of the two comparison objects was longer or heavier in the length and weight tasks, respectively. With this procedure, the results of the previous studies were confirmed: Memory for premises and transitivity judgments were statistically unrelated. Brainerd and Kingma explained these results in terms of a "fuzzy trace theory," according to which children do not solve transitivity problems by coordinating precise memory traces of the individual premise relations, "A is longer than B," is longer than C," and so on. Instead, they solve such problems with reference to imprecise (i.e., "fuzzy") memory traces of the form, "Things get bigger toward the right." Memory for individual premise relations are stochastically independent of transitivity judgments, because those judgments are obtained by other means.

This "fuzzy trace theory" was advanced by Brainerd and Kingma (1984) as a general theory of transitive reasoning (indeed, as a theory of children's reasoning in general—see Brainerd & Kingma, 1985). As mentioned previously, however, Chapman and Lindenberger (1988) argued that children could solve transitivity tasks through the use of right-left spatial position cues only if the relevant dimension (length or weight) was in fact correlated with spatial position (as in their "alternate" transitivity tasks). If, in contrast, the relevant dimension was uncorrelated with position cues (as in their "standard" tasks), children would be unable to use position cues and would solve the transitivity tasks through a composition of premise relations instead. Dependencies between children's transitivity judgments and their memory for premises would be expected only under the "standard" conditions. Because the procedures used by Brainerd and Kingma in assessing transitive reasoning were comparable to Chapman and Lindenberger's "alternate" task, the absence of reasoning-remembering dependencies in Brainerd and Kingma's studies of transitivity can be explained by the fact that their task did not require the form of reasoning for which reasoning-remembering dependencies would be expected. As indicated in the present analyses, reasoning-remembering dependencies indeed are found when children are tested under the "standard" conditions instead.

The same principle holds for the other cognitive tasks for which Brainerd and Kingma (1984, 1985) reported a reasoning-remembering independence. A dependency between judgments and memory for premises would be expected only under conditions in which children actually refer to the specific premises for which memory is tested in making their judgments. For example, Brainerd and Kingma tested the relation between children's class inclusion judgments and their memory for the cardinal

numbers of objects in the relevant superordinate and subordinate classes. Although the knowledge that there are 7 boys and 3 girls and 10 children altogether might be *sufficient* for inferring that "there are more children than boys," it is not *necessary* for such a judgment. According to Inhelder and Piaget's (1959/1969) further analysis, the class inclusion problem involves the quantification of classes with the logical quantifiers "all" and "some." Thus, a typical justification of a correct class inclusion judgment has the following form: "There are more children than boys, *because they're all children but only some are boys.*" Another typical justification has the form of class addition: ". . . because the boys and girls are both children." Such forms of reasoning could be employed in inferring class inclusion judgments, even if children did not remember the exact cardinality of the classes involved. Because memory for the cardinality of classes is not a necessary condition for inferring a correct judgment with concrete operational reasoning, a dependency between memory for cardinality and class inclusion judgments is not to be expected.

The argument is similar for Brainerd and Kingma's (1985) probability judgment task. In this case, the premises for which memory was tested were the cardinal numbers of two types of animals (e.g., 7 turtles and 3 birds) in an opaque container, and the question was which animal was most likely to be drawn by chance. But children could have inferred that a turtle was *more likely* from the fact that the container contained *more* turtles, quite apart from their exact number. In other words, children could have inferred likelihood as a function of *relative* numerosity without necessarily remembering the cardinal numbers of each animal.

In contrast, Brainerd and Kingma (1984) did test the relation between conservation judgments and memory for those premises that children actually might be expected to use in inferring those judgments. According to Piaget and Inhelder's (1941/1974) analysis, children solve the familiar conservation of liquid problem in the following way: From the knowledge that (a) the amount of liquid in beaker *A* is the same as that in beaker *B* of the same size and shape, together with the knowledge that (b) the amount of liquid in *B* remains the same when it is poured into beaker *B'* of a different shape, children can infer the judgment that (c) the amount of liquid in *A* is the same as that in *B'*. Brainerd and Kingma (1984) found a significant dependence of conservation judgments ( $A = B'$ ) on memory for one of those premises ( $B = B'$ ), but not for the other one ( $A = B$ ). According to the distributional model described in this paper, such a result could be explained by the fact that the probability of accurate memory for  $B = B'$  was relatively low (0.353), but that for  $A = B$  was quite high (0.801). Given  $N = 95$  and the probability of a correct judgment equal to 0.249 (Brainerd & Kingma, 1984, p. 369), Eq. 1 yields a value of 0.078 for  $B = B'$  and 0.189 for  $A = B$ . In other words, 8% or more of the sample would have to be operational reasoners in order to obtain a sig-

nificant dependence between the premise  $B = B'$  and the conservation judgment  $A = B'$ , but about 19% or more would be necessary to obtain dependence between  $A = B$  and the same conservation judgment. Thus, an absence of a dependency between conservation judgments and the presumed premise  $A = B$  does not necessarily imply that children did not employ that premise in inferring their judgments. Instead, the data are consistent with the conclusion that Brainerd and Kingma's sample comprised between 8 and 19% operational reasoners who used both premises. In a sample of kindergartners and first graders, such a range is not unlikely.

The overall point is that, even if the major condition for testing reasoning-remembering dependencies is met—that memory is tested for the premises actually used under the particular task conditions—the present analysis indicates that such reasoning-remembering dependencies would be found only if the proportion of operational reasoners in the sample is sufficiently large. The value of the present distributional model is that it provides a way of estimating this critical proportion, given the probabilities of accurate memory and correct judgments. The critical proportion of operational reasoners is higher, the higher the probabilities of accurate memory and correct judgment.

More generally, the analysis of the statistical dependencies between judgments and memory for premises provides the researcher with a general method for testing hypotheses regarding the forms of reasoning by which those judgments were inferred. The hypothesis that a given judgment,  $J$ , is derived inferentially from a given set of premises,  $P$ , can be tested by examining the statistical dependency between  $J$  and memory for  $P$ —provided that the minimal conditions for the use of this method as described in this article are fulfilled. One important advantage of this method is that it can be applied independently of verbal justifications, which previously have been the most frequently used means of distinguishing among differing structural models of inferential reasoning. The major problem with using children's *verbal* justifications for assessing the forms of their *intrapsychic* inferences is that a justification of a judgment once obtained may or may not reflect the process by which that judgment was generated. As Nisbett and Wilson (1977) and others have found for adult subjects, justifications might be after-the-fact "rationalizations" of conclusions arrived at by other means. As a result of these methodological considerations, the trend in cognitive-developmental psychology has been to ignore children's verbal justifications more or less entirely and to focus instead on judgments only or on various "nonverbal" assessment procedures (Braine, 1959; Brainerd, 1973; Siegel, 1978). Unfortunately, the results obtained through such procedures have been subject to interpretive ambiguities of their own (Larsen, 1977; Miller, 1976).

As stated in the introduction to this article, our position is that the

exclusive focus on intrapsychic inference in cognitive-developmental psychology has led to a neglect of verbal justifications as phenomena of interest in their own right (Chapman, 1991). One reason why justifications are of interest is because, according to some major theorists (Piaget, 1924/1928; Vygotsky, 1934/1986), intrapsychic inference develops through *sociogenesis*—more precisely through the interiorization of interpersonal argumentation. The significance of verbal justifications in this connection is that they are the primary moves in argumentation; one attempts to convince other persons of one's own point of view by providing them with reasons that would justify its acceptance. Moreover, if intrapsychic inference develops from an interpersonal context of justification, then one would expect some congruence between the development of competence in justification and the development of competence in intrapsychic inference. Once a particular form of reasoning is mastered, it could be used both for purposes of justification (in moving from a given conclusion to the premises which justify it) and for deductive inference (in moving from a given set of premises to the conclusion's) which follow from them. In this view, justification and intrapsychic inference are conceptually distinct, but theoretically related competencies. The nature and degree of the relation between them are empirical questions.

Those questions can be addressed with the method outlined in this article. If children use the same form of reasoning for justifying their judgments as they used in inferring those judgments, then children should justify their judgments with reference to the same premises from which they originally derived those judgments and a reasoning-remembering dependency between judgments and memory for said premises should be found. In effect, this hypothesis was tested in the analyses shown in Tables 4 and 5. The results were straightforward: Children giving operational justifications (" $A > C$ , because  $A > B$  and  $B > C$ ") showed precisely the pattern of reasoning-remembering dependency to be expected if they also had inferred their judgments through a composition of the same premises (i.e., " $A > B$  and  $B > C$ , therefore  $A > C$ ").

However, such results should *not* be interpreted to the effect that justifications are valid measures of inference after all. For one thing, the problem of false negatives remains; an unknown number of children who gave nonoperational justifications nevertheless might have inferred their judgments from a composition of  $A > B$  and  $B > C$ . For another thing, one cannot assume that such a congruence between justifications and inferences would be found for other tasks or under other task conditions. In our view, the problem with previous discussions of children's verbal justifications as response criteria was the assumption that justifications are of interest only as potential measures of intrapsychic inference. Instead,

we suggest that justification should be considered a distinct competence of theoretical interest in itself and that the relation *between* justification and intrapsychic inference is an important question for theory and research.

### THEORIES OF TRANSITIVE INFERENCE

In this final section, the broader implications of the foregoing results for theories of children's reasoning are discussed. Four different theories are briefly considered: (a) the *fuzzy trace* theory of Brainerd and his colleagues (Brainerd & Kingma, 1984, 1985; Brainerd & Reyna, 1989, 1990; Reyna & Brainerd, 1990), (b) the *mental logic* approach, typified by Braine and Rumain (1983), (c) the *mental models* approach of Johnson-Laird (1983), and (d) the *operational-constructive* theory, which we derive from Piaget. We begin with fuzzy trace theory, because it provided the original impetus for this investigation and because it differs from the other three in its emphasis on nondeductive processes in reasoning. We then continue with an attempt to differentiate the other three approaches as theories of deductive inference, with particular emphasis on developing a nontraditional version of Piagetian constructivism.

#### *Fuzzy Traces versus Deductive Inference*

In Brainerd and Reyna's (1990) comprehensive statement of fuzzy trace theory to date, the theory was summarized in 7 points. The last 3 of those points, dealing with issues of response interference, cognitive resources, and development, respectively, lie outside the immediate scope of this discussion (see Brainerd & Reyna, 1989, and the accompanying commentaries). Instead, we limit ourselves here to a consideration of the first 4 points insofar as they are relevant to theories of children's reasoning and to theories of transitive reasoning in particular:

1. At all ages, children encode incoming information not only in terms of direct, "verbatim" traces of that information, but also in terms of the basic "pattern" or "gist" that can be extracted from the original information.
2. As a result, the information represented in working memory varies along a continuum of exactness, from verbatim traces at one end to fuzzy traces at the other. This continuum makes possible certain processing options in reasoning and other tasks, depending on which type of traces are being processed. However:
3. Choice among possible options is governed by a preference for fuzzy processing because, among other reasons, they are more accessible, less complex, and require less effort.
4. Such processing options extend to short-term memory. That is, short-term memory probes can be answered through either the simple

retrieval of memory traces (fuzzy or verbatim), or through their reconstruction.

This last-named feature of the theory was said to account both for the occurrence of reasoning-remembering dependencies when they occur and for their absence when they do not occur. Briefly, the argument was that such dependencies should occur when memory probes can be answered through reconstructive processing of the same level of gist used in reasoning, because then there is complete overlap between reasoning and remembering with respect to both the level of traces (fuzzy or verbatim) and the type of processing (reconstructive). In contrast, reasoning-remembering *independence* would be expected if memory probes are answered through simple retrieval of either kind of trace, because processing operations (and/or level of gist) are different for memory probes and reasoning tasks (Brainerd & Reyna, 1990, p. 33).

The fuzzy trace theory of transitive reasoning was explained by Brainerd and Reyna (1990, pp. 11–12) as follows: Children are presented with certain “background facts,” for example, three rods *A*, *B*, and *C*, of lengths 18, 17.5, and 17 cm, respectively. Children code these background facts along the continuum of exactness, from verbatim traces that reflect the background facts more or less exactly to fuzzy traces having the form, “Things get longer to my left.” (Between these two extremes are traces of intermediate exactness having the form, “*A* is long,” “*B* is short,” etc.). As previously described, this continuum allows for certain processing options: “Transitivity problems can be solved with left-right discrimination operations if global spatial patterns (“things get longer to my left”) are being processed, but logical operations may be required if the verbatim premises are being processed” (Brainerd & Reyna, 1990, p. 18). Brainerd and Reyna cited Chapman and Lindenberg (1988) for an example of such processing options, implying that the different results obtained by Chapman and Lindenberg in their “alternate” and “standard” transitivity tasks can be explained in this way.

In our opinion, some deep ambiguities exist in fuzzy trace theory as summarized in the preceding paragraphs. First, there is ambiguity about the nature of “verbatim traces” and their relation to the “background facts” that they are supposed to represent. Brainerd and Reyna give the “background facts” for the transitivity task in terms of the exact lengths of the rods or sticks to be compared. The notion of “verbatim traces” seems to imply that children form direct percept-like “copies” of those lengths. However, one could state the “background facts” alternately in terms of the *relations* among comparison objects ( $A > B$ , etc.) or even (task conditions permitting) in terms of the spatial arrangement (“Things get bigger to the left”). What the “background facts” are in any given task is not an unequivocal proposition, and one wonders why only the

more exact representations of the facts (whether in terms of exact lengths or length relations) qualify as "verbatim traces." We would suggest that the transitivity task is usually presented through the use of a verbal comparative (e.g., "longer") as to maximize the probability that children represent the problem in terms of length *relations*. "Verbatim" representations could be construed, as the etymology of the term implies, in terms of the *words* employed. Thus, a relational representation (e.g.,  $A > B$ ) could be considered "verbatim" in a task in which the comparative "longer" was used explicitly in the presentation of premise information.

The idea that the facts, however they are stated, can be represented in alternative ways by children goes without question. But we think that the best way of conceptualizing these alternatives is not necessarily in terms of the fuzzy-to-verbatim continuum. To us, the "verbatim" representations and the "fuzzy" traces described by Brainerd and Reyna appear to differ *in form*, not merely in exactness. Representations of the form  $A = 18$  cm and  $B = 17.5$  cm are *numerical*; representations of the form, "A is longer than B," are *relational*; representations of the form, "A is long," and "B is short," are *categorical*; and representations of the form, "things get bigger to the right," can be called *functional* in the sense that relations of size ("bigger") are represented as a function of spatial relations ("to the right"). Presumably, any of these forms of representation could be more or less exact. Indeed, formal systems of "fuzzy reasoning" have been devised in which each type of term is allowed to vary in "fuzziness" (e.g., fuzzy sets, fuzzy relations, fuzzy functions, fuzzy numerical information—see Zedah, 1975). In fuzzy trace theory, the *form* of representation is confounded with *fuzziness* in this sense.

More problematic in our opinion are the claims regarding the scope of fuzzy trace theory and the explanation of reasoning-remembering dependencies. In the passage quoted above, Brainerd and Reyna (1990) stated that the processing of fuzzy traces occurs when "global spatial patterns" permit; otherwise logical operations may be necessary. If so, then reasoning by fuzzy traces occurs only under special conditions, and the problem of explaining how children reason when those conditions are *not* met remains. This was Chapman and Lindenberger's (1988) argument: that solving a transitivity task by means of a "fuzzy trace" relating length or weight to spatial orientation is possible only under the special conditions that length or weight is actually correlated with spatial position.

With respect to the explanation of reasoning-remembering dependencies, we believe that fuzzy trace theory is not entirely consistent. Brainerd and Reyna (1990) argue that reasoning-remembering dependencies should be found when reasoning tasks and memory probes can be solved in the same way, that is, with the same traces and processes. This condition would appear to be met in the case of transitive reasoning when the



relevant dimension is correlated with spatial position (i.e., in Brainerd & Kingma's, 1989, studies or in Chapman & Lindemberger's "alternate" transitivity task). In that case, a fuzzy trace of the form, "Things get longer to my left," should be sufficient to answer the transitivity question regarding the nonadjacent  $A-C$  comparison as well as the memory probes regarding the adjacent  $A-B$  and  $B-C$  comparisons through reconstructive processing. However, fuzzy trace processing was cited by Brainerd and Kingma (1984) as a reason why transitivity judgments and memory for nonadjacent premise comparisons ought *not* to be found on this version of the transitivity task, and indeed, they found no evidence for reasoning-remembering dependency in a number of studies.

According to Reyna and Brainerd (1990), reasoning-remembering independence occurred in the Brainerd and Kingma studies because of errors of retrieval, of misunderstanding, of misperception, or of misidentification. This explanation seems eminently reasonable to us, but it would apply as well to all other cases in which memory probes can be answered in the same way as reasoning problems—that is, even in those cases which, according to Brainerd and Reyna (1990), fuzzy trace theory would predict reasoning-remembering *dependence*. We conclude that fuzzy trace theory is not entirely consistent on this point.

One further problem is worth mentioning: Even if fuzzy trace theory accurately describes how children reason under some conditions, it does not address the problems of how and why some forms of reasoning are accompanied by a "sense of necessity"—the feeling that a given conclusion simply could not be otherwise. An inference of the form, "A is longer than B, because A is to the left of B," may be objectively correct under the special conditions that length is correlated with right-left spatial position, but the correctness of this inference depends on situation-specific empirical knowledge of the correlation between length and position. In contrast, both children and adults recognize an element of necessity in transitive inference (Miller, 1986), and explaining this sense of necessity and its development is a major task of an epistemologically sophisticated psychology of reasoning (Bickhard, 1988; Piaget, 1970/1972b).

In summary, we believe that there are a number of ambiguities and/or inconsistencies in fuzzy trace theory which prevent it from being a satisfactory theory of children's reasoning. At best, we think that it might be sufficient for explaining how children and adults might reason under certain special conditions. But one must still explain how individuals reason under more generally and, in particular, how they reason according to standard deductive logic when in fact they do so.

### *Mental Logic versus Mental Models*

Johnson-Laird (1983) contrasted two general approaches to human reasoning which he labeled "mental logic" and "mental models." In con-

trast to fuzzy trace theory, both of these approaches are explicit theories of deductive reasoning. "Mental logic" is the view that individuals draw inferences in reasoning tasks by applying rules of logic. The major problem in this approach is to determine what these rules actually are. For Johnson-Laird, Piaget was the chief proponent of the mental logic, but Braine and Rumin's (1983) effort to identify a natural logic of inference schemas based on theorems of propositional logic could be said to typify this approach as well. (The question of whether Piaget is appropriately classed as an advocate of mental logic is discussed later.)

Curiously, Braine and Rumin do not discuss transitive reasoning at all, preferring to consider the knowledge of the transitivity inherent in relations of length or weight as empirical, rather than "logical," in nature. This dismissal of the "logical" character of transitive reasoning might be considered a shortcoming in itself, considering the fact that Miller's (1986) subjects classed transitive reasoning with analytical deduction as involving a sense of "logical" necessity. In any case, one can easily imagine how one might attempt to explain transitive reasoning within the mental logic approach. Briefly, one might assume that, given the premise comparisons  $A > B$  and  $B > C$ , children apply a general "rule of transitivity." Such a rule might be represented mentally in different ways, but stated in English it would have something like the following form: "For every length  $X$ ,  $Y$ , and  $Z$ , if  $X > Y$  and  $Y > Z$ , then  $X > Z$ ." The transitivity problem would be solved by applying this rule to the problem at hand—that is, by evaluating the variables  $X$ ,  $Y$ , and  $Z$  in terms of the objects  $A$ ,  $B$ , and  $C$ , respectively. Such an application returns the conclusion  $A > C$ .

Although the foregoing model provides an explanation of how transitive inferences are carried out (by applying the transitivity rule), it does so at the expense of begging the question of how the rule and the sense of necessity are acquired. Presumably, subjects understand the necessity of transitive reasoning, because that necessity is embodied in the rule. But the question remains how and understanding of the rule with its attendant sense of necessity develops originally. In short, the mental logic approach is especially vulnerable to the "learning paradox" as described by Pascual-Leone (1987): The understanding of transitivity is explained in a way that presupposes the understanding to be explained.

Johnson-Laird's (1983) criticism of the mental logic approach was somewhat different. At best, mental logic can explain how individuals reason correctly. It does not explain the kinds of logical errors in reasoning that are typically observed in adult subjects, nor the kinds of extralogical factors that influence the difficulty of different reasoning problems. His "mental models" approach was an attempt to address such problems. In his view, subjects use premise information to construct schematic "tableaus" (or "mental models") which represent the problem, and those representations already contain the information pertaining

to the solution. Subjects obtain the solution simply by decoding the information contained in the models. For example, the transitivity of length problem is solved, not by applying a rule of transitivity, but by mentally aligning the objects  $A > B > C$  along their length axes. Such a construction makes the fact that  $A > C$  obvious to intuition. By applying this approach to a number of different tasks, Johnson-Laird was able not only to explain how subjects normally solved them, but also to predict the ways in which a number of extralogical factors influenced their performance.

Because of Johnson-Laird's success in providing explanations for the errors and biases typically found in everyday reasoning, Galotti (1989) named the mental models approach as the most promising approach to a general theory of human reasoning. Nevertheless, its reliance on mental "tableaus" makes it vulnerable to the epistemological problems inherent in representational theories of mind, such as how the subject knows that a given tableau is an accurate representation of reality (Russell, 1987). Moreover, one wonders how to explain the sense of necessity inherent in transitive reasoning within the mental models approach. Like most other psychological theories of deductive reasoning, it is an attempt to explain how subjects derive correct inferences, not how they are able to know whether or not their inferences are *necessarily* correct. In short, the mental models approach, like the mental logic approach, begs some important epistemological questions.

### *Operational Constructivism*

As mentioned previously, Piaget's theory of reasoning is commonly interpreted as an example of the mental logic approach, that is, as implying that children solve reasoning problems through the application of logical rules as specified in his system of operator logic (Brainerd & Reyna, 1990; Johnson-Laird, 1983; Pascual-Leone, 1987). We believe that this interpretation misconstrues the constructivist thrust of his epistemological theory and propose an alternative interpretation. The point is not necessarily that our interpretation reflects "what Piaget really believed." Instead, we attempt to reconstruct the sense of operator logic in the manner that we believe most consistently coheres with a constructivist epistemology.

In this view, operator logic is not a system of logical rules to be applied to a problem in order to infer a solution. Instead, it is a structural description of the system of operations involved in comprehending the problem. Thus, children comprehend physical relations involving length and weight in terms of certain *operational criteria*. For example, the relation "longer than" is comprehended in terms of an implicit operation of comparison in which one aligns the objects to be compared along their

length axes and notes which object protrudes beyond the other. That is what it *means* for one object to be "longer than" another. It was in this sense that Piaget interpreted physical relations in terms of differences (e.g., Inhelder & Piaget, 1959/1969, p. 11). This approach does not entail a "reduction" of relations to differences (Seltman & Seltman, 1985); rather the criteria for comprehending the relation are the operations involved in noting a difference. (On the Wittgensteinian concept of "criteria," see Chapman, 1987b.)

The operational constructivist model of transitive reasoning can be understood in this context. Grasping the relation "*A* is longer than *B*" involves an implicit cognitive operation in which the two objects are aligned such that the *difference* in length between *A* and *B* is affirmed. An analogous operation is involved in grasping the relation between *B* and *C*. These operations of comparison play the role of the "first-order operations" discussed previously with respect to Piaget's theory of concrete operations. When asked about the relation between *A* and *C*, children apply a second-order operation to those first-order operations. Whereas the first-order operations generate differences in length between *A* and *B* and between *B* and *C*, the second order operation *adds* those differences together to obtain the total difference between *A* and *C* (cf. Inhelder & Piaget, 1959/1969, p. 11). For this purpose, precise estimates of the magnitudes of the *AB* and *BC* differences are unnecessary. The knowledge that one positive quantity (e.g.,  $A - B$ ) added to another ( $B - C$ ) yields a third positive quantity ( $A - C$ ) is sufficient to infer the fact that *A* is longer than *C* (i.e., that the difference  $A - C$  is positive).

With this model of transitive reasoning, one can explain both the content-dependence of the transitivity of length as well as the sense of "logical" necessity associated with it. On the one hand, both the first-order operations of comparison as well as the second-order addition of lengths are specifically adapted to the physical dimension of length. Presumably, such adaptedness results from the fact that both the first- and second-order operations develop from the interiorization of physical actions bearing on length. On the other hand, compositions of the first-order operations have uniquely determined results which are invariant across other physical dimensions such as color or shape. That uniqueness and invariance is the source of the sense of necessity—the sense that, given the outcomes of two comparisons of length (e.g., *AB* and *BC*), the outcome of a third comparison (*AC*) is not only fully determined, but *could not be otherwise*.

By identifying deductive reasoning as an important phenomenon to be explained, we do not thereby commit ourselves to the view that rational thinking is essentially deductive in character, nor that deductive reasoning, once it develops, replaces all other forms of reasoning. On the con-

trary, we agree with Brainerd and Reyna (1990) that children and adults possess optional forms of reasoning for solving particular problems and that, other things being equal, certain forms of reasoning will be preferred to others according to a principle of cognitive economy. In particular, we have argued that much of what passes for preoperational thinking in Piagetian theory can be described in terms of *functional inference*: deriving one dimension of experience as a function of another (Piaget, Grize, Szeminska, & Vinh Bang, 1967/1977). That capacity does not disappear with the development of deductive thought. Instead, it continues to exist in an enriched form as an alternative form of reasoning which has both advantages and disadvantages with respect to deduction. On the positive side, functional reasoning generally requires less mental effort than deduction. In contrast, it lacks the certainty and necessity of deductive inference and is applicable only under a limited range of conditions when the dimensions in question are in fact functionally related. With respect to transitive reasoning, our argument has been that functional reasoning can only be employed to solve transitivity problems under the specific condition that relations of length are objectively correlated with spatial orientation. When those special conditions do not obtain, the analyses presented in this article suggest that functional reasoning cannot be employed and that deductive solutions are employed instead.

### *Conclusion*

The results reported in this article indicate that, under suitable task conditions which prevent alternate solution strategies, a dependency exists between transitivity judgments and memory for premise relations. This finding is consistent with the conclusion that, under the stated conditions, children solve transitivity problems by deduction from pairwise premise comparisons rather than from "fuzzy traces" of the type suggested by Brainerd and Kingma (1984). The main point was that testing for reasoning-remembering dependencies is a general method that can be used for determining whether a given conclusion in fact was derived from a given set of premises. A question for further research is whether it might also prove useful for discriminating among the different *theories* of deductive inference as described above. In any case, the general usefulness of the method depends on knowing the conditions under which one can expect it to yield valid conclusions. Describing those conditions was the main task of this article.

### REFERENCES

- Bickhard, M. (1988). Piagetian variation and selection models: Structuralism, logical necessity, and interactivism. *Human Development*, 31, 274-312.
- Braine, M. D. S. (1959). The ontogeny of certain logical operations: Piaget's formulation

- examined by nonverbal methods. *Psychological Monographs: General and Applied*, 73(5), Whole No. 475), 1-43.
- Braine, M. D. S., & Rumain, B. (1983). Logical reasoning. In P. H. Mussen (Ed.), *Handbook of child psychology, Vol. 3: Cognitive development* (pp. 263-340). New York: Wiley.
- Brainerd, C. J. (1973). Judgments and explanations as criteria for the presence of cognitive structures. *Psychological Bulletin*, 79, 172-179.
- Brainerd, C. J. (1977). Response criteria in concept development. *Child Development*, 48, 360-366.
- Brainerd, C., & Kingma, J. (1984). Do children have to remember to reason? A fuzzy-trace theory of transitivity development. *Developmental Review*, 4, 311-377.
- Brainerd, C., & Kingma, J. (1985). On the independence of short-term memory and working memory in cognitive development. *Cognitive Psychology*, 17, 210-247.
- Brainerd, C. J., & Reyna, V. F. (1989). Output-interference theory of dual-task deficits in memory development. *Journal of Experimental Child Psychology*, 47, 1-18.
- Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, 10, 3-47.
- Breslow, L. (1981). Reevaluation of the literature on the development of transitive inferences. *Psychological Bulletin*, 89, 325-351.
- Bryant, P. E., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature*, 232, 456-458.
- Chapman, M. (1987a). Piaget, attentional capacity, and the functional implications of formal structure. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 20, pp. 289-334). Orlando, FL: Academic Press.
- Chapman, M. (1987b). Inner processes and outward criteria: Wittgenstein's importance for psychology. In M. Chapman & R. A. Dixon (Eds.), *Meaning and the growth of understanding* (pp. 103-127). Berlin: Springer-Verlag.
- Chapman, M. (1988). *Constructive evolution: Origins and development of Piaget's thought*. Cambridge: Cambridge University Press.
- Chapman, M. (1989). Resources versus response competition: A false disjunction? *Journal of Experimental Child Psychology*, 47, 39-41.
- Chapman, M. (1991). The epistemic triangle: Operative and communicative components of cognitive competence. In M. J. Chandler and H. M. Chapman (Eds.), *Criteria for competence: Controversies in the conceptualization and assessment of children's abilities* (pp. 209-228). Hillsdale, NJ: Erlbaum.
- Chapman, M. (in press). Everyday reasoning and the revision of belief. In J. Puckett & H. W. Reese (Eds.), *Life-span developmental psychology: Mechanisms of everyday cognition*. Hillsdale, NJ: Erlbaum.
- Chapman, M., & Lindenberger, U. (1988). Functions, operations, and decalage in the development of transitivity. *Developmental Psychology*, 24, 542-551.
- Chapman, M., & Lindenberger, U. (1989). Concrete operations and attentional capacity. *Journal of Experimental Child Psychology*, 47, 236-258.
- Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *The Behavioral and Brain Sciences*, 4, 317-370.
- De Boysson-Barides, B., & O'Regan, K. (1973). What children do in spite of adults' hypotheses. *Nature*, 246, 531-534.
- Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin*, 105, 331-351.
- Gruen, G. E. (1966). Note on conservation: Methodological and definitional considerations. *Child Development*, 37, 977-983.

- Halford, G. S., & Galloway, W. (1977). Children who fail to make transitive inferences can remember comparisons. *Australian Journal of Psychology*, *29*, 1-5.
- Halford, G. S., & Kelly, M. E. (1984). On the basis of early transitivity judgments. *Journal of Experimental Child Psychology*, *38*, 42-63.
- Hall, V. C., & Kaye, D. B. (1978). The necessity of logical necessity in Piaget's theory. In L. S. Siegel & C. J. Brainerd (Eds.), *Alternatives to Piaget* (pp. 153-167). New York: Academic Press.
- Hildebrand, D. K., Laing, J. D., & Rosenthal, H. (1977). *Prediction analysis of cross classifications*. New York: Wiley.
- Howe, M. L., & Rabinowitz, F. M. (1990). Resource panacea? Or just another day in the developmental forest? *Developmental Review*, *10*, 125-154.
- Inhelder, B., & Piaget, J. (1969). *The early growth of logic in the child*. New York: Norton (Original work published 1959)
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- Kallio, K. (1982). Developmental change on a five-term transitive inference. *Journal of Experimental Child Psychology*, *33*, 142-164.
- Kingma, J. (1981). *De ontwikkeling van quantitative en relationele begrippen bij kinderen van 4 tot 12 jaar*. Unpublished doctoral dissertation, State University of Groningen, The Netherlands.
- Larsen, G. Y. (1977). Methodology in developmental psychology: An examination of research on Piagetian theory. *Child Development*, *48*, 1160-1166.
- Miller, S. A. (1976). Nonverbal assessment of Piagetian concepts. *Psychological Bulletin*, *83*, 405-430.
- Miller, S. A. (1986). Certainty and necessity in the understanding of Piagetian concepts. *Developmental Psychology*, *22*, 3-18.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231-259.
- Pascual-Leone, J. (1987). Organismic processes for neo-Piagetian theories: A dialectical causal account of cognitive development. *International Journal of Psychology*, *22*, 531-570.
- Perner, J., & Aebi, J. (1985). Feedback-dependent encoding of length series. *British Journal of Developmental Psychology*, *3*, 133-141.
- Piaget, J. (1928). *Judgment and reasoning in the child*. London: Routledge & Kegan Paul. (Original work published 1924)
- Piaget, J. (1950). *The psychology of intelligence*. New York: Harcourt Brace. (Original work published 1947)
- Piaget, J. (1955). *The language and thought of the child*. Cleveland: Meridian. (Original work published 1923)
- Piaget, J. (1971). The theory of stages in cognitive development. In D. R. Green (Ed.), *Measurement and Piaget* (pp. 1-11). New York: McGraw-Hill.
- Piaget, J. (1972a). *Essai de logique opératoire*. Paris: Dunod. (2nd ed. of *Traité de logique*, 1949)
- Piaget, J. (1972b). *The principles of genetic epistemology*. London: Routledge & Kegan Paul. (Original work published 1970)
- Piaget, J. (1986). Essay on necessity. *Human Development*, *29*, 301-314. (Original work published 1977)
- Piaget, J. (1987a). *Possibility and necessity. Vol. 2: The role of necessity in cognitive development*. Minneapolis: University of Minnesota Press. (Original work published 1983)
- Piaget, J. (1987b). *Possibility and necessity. Vol. 1: The role of possibility in cognitive development*. Minneapolis: University of Minnesota Press. (Original work published 1981)

- Piaget, J., Grize, J.-B., Szeminska, A., & Vinh Bang. (1977). *Epistemology and psychology of functions*. Dordrecht, Holland: Reidel. (Original work published 1967)
- Piaget, J., & Inhelder, B. (1974). *The child's construction of quantities*. London: Routledge & Kegan Paul. (Original work published 1941).
- Reese, H. W., & Schack, M. L. (1974). Comment on Brainerd's criteria for cognitive structures. *Psychological Bulletin*, *81*, 67-69.
- Rescher, N. (1988). *Rationality*. Oxford: Clarendon Press.
- Reyna, V. F., & Brainerd, C. J. (1989). Output interference, generic resources, and cognitive development. *Journal of Experimental Child Psychology*, *31*, 42-46.
- Reyna, V. F., & Brainerd, C. J. (1990). Fuzzy processing in transitivity development. *Annals of Operations Research*, *23*, 37-63.
- Roodin, M. L., & Gruen, G. E. (1970). The role of memory in making transitive judgments. *Journal of Experimental Child Psychology*, *10*, 264-275.
- Russell, J. (1981). Children's memory for the premises in a transitive measurement task assessed by elicited and spontaneous justifications. *Journal of Experimental Child Psychology*, *31*, 300-309.
- Russell, J. (1987). Rule-following, mental models, and the developmental view. In M. Chapman & R. A. Dixon (Eds.), *Meaning and the growth of understanding* (pp. 23-48). Berlin: Springer-Verlag.
- Scribner, S. (1977). Modes of thinking and ways of speaking: Culture and logic reconsidered. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking* (pp. 482-500). Cambridge, England: Cambridge University Press.
- Seltman, M., & Seltman, P. (1985). *Piaget's logic*. London: Allen & Unwin.
- Siegel, L. S. (1978). The relationship of language and thought in the preoperational child: A reconsideration of nonverbal alternatives to Piagetian tasks. In L. S. Siegel & C. J. Brainerd (Eds.), *Alternatives to Piaget* (pp. 43-67). New York: Academic Press.
- Trabasso, T. (1975). Representation, memory, and reasoning: How do we make transitive inferences? In A. D. Pick (Ed.), *Minnesota Symposium on Child Psychology* (Vol. 9, pp. 135-172). Minneapolis: University of Minnesota Press.
- Trabasso, T. (1977). The role of memory as a system in making transitive inferences. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 333-366). Hillsdale, NJ: Erlbaum.
- von Eye, A., & Brandtstädter, J. (1988). Evaluating developmental hypotheses using statement calculus and nonparametric statistics. In P. B. Baltes, D. Featherman, & R. M. Lerner (Eds.), *Life-span development and behavior* (Vol. 8, pp. 61-97). Hillsdale, NJ: Erlbaum.
- Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: MIT Press. (Original work published 1934)
- Wason, P. C., & Johnson-Laird, P. M. (1972). *Psychology of reasoning*. London: Batsford.
- Youniss, J., & Furth, H. (1973). Reasoning and Piaget. *Nature*, *244*, 314-315.
- Zedah, L. A. (1975). Fuzzy logic and approximate reasoning. *Synthese*, *30*, 407-428.

RECEIVED: August 1, 1989; REVISED: August 27, 1990.