DOCUMENT RESUME

ED 097 977 PS 007 513

AUTHOR Brainerd, Charles J.
TITLE On the Consequences of Type I and Type II Criterion Errors in Concept Development Research.
PUB DATE [74]
NOTE 26p.

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS Abstraction Levels; Age Groups; *Cognitive Measurement; *Concept Formation; *Developmental Psychology; Psychometrics; *Research Criteria; Research Design; *Sequential Learning; Training Objectives
IDENTIFIERS *Piaget (Jean)

ABSTRACT

The criterion problem in the neoPiagetian cognitive-developmental literature is reviewed. This problem has provoked debate in the context of three empirical questions: concept age norms, concept training effects, and concept invariant sequences. It is argued that only the question of invaniant sequences in same-stage concepts is theoretically crucial. With regard to this question it is shown that both Type I ("false positive") and Type II ("false negative") criterion errors tend to produce data which are spuriously supportive of theoretical predictions. It is concluded that the objective psychometric consequences of criterion errors provide no grounds for the current view that obtaining or not obtaining theory-supportive findings in concept development research turns on arbitrary decisions about response criteria. (Author/CS)

ERIC

ED 097977

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

AUG. 28 1974

ON THE CONSEQUENCES OF TYPE I AND TYPE II CRITERION ERRORS IN CONCEPT DEVELOPMENT RESEARCH

Charles J. Brainerd

University of Alberta

PS 007513

This paper is concerned with a debate over the response criteria used to infer the presence of Piagetian stage-related concepts which has been going on in the cognitive-developmental literature for more than a decade. The debate began during the early 1960s with a series of exchanges between Braine (1962, 1964) and Smedslund (1963, 1965) over transitivity criteria. Shortly thereafter, the debate was generalized to the measurement of other Piagetian concepts in papers by Gruen (1966) and Smedslund (1969). The present decade has witnessed a continuation of the original disagreement over transitivity criteria in the form of an exchange between Trabasso and his associates (Bryant, 1973; Bryant & Trabasso, 1971; Riley & Trabasso 1974) and Youniss and Furth (1973). Further discussions of the general version of the criterion problem, which includes the transitivity disagreement as a special case, also have appeared recently (Brainerd, 1973a, 1973b, 1973c, 1974a; Brainerd & Hooper, 1974; Kuhn, 1974; Reese & Schack, 1974).

In its general form, the criterion problem is concerned with the minimum behavioral evidence an investigator requires before he will conclude that a given subject possesses some stage-related concept. On the one hand, the Genevans (e.g., Inhelder, Bovet, Sinclair, & Smock; Inhelder & Sinclair, 1969; Sinclair, 1973) and investigators who favor a more or less orthodox Piagetian view of cognitive development (e.g., Lasry & Laurendeau, 1969; Smedslund, 1963; Strauss, 1972; Youniss & Furth, 1973) require very strong evidence. On the other hand, less orthodox investigators (e.g., Braine, 1959; Braine & Shanks, 1965a, 1965b, Bruner, 1964; Bryant & Trabasso, 1971; Mehler & Bever, 1967) will accept somewhat weaker evidence. For reasons that have never been explicitly stated, investigators from the former group seem to view Type I assessment errors ("false positive" diagnoses) as far more heinous than Type II assessment errors ("false negative" diagnoses)--hence, their conservative criteria. In contrast, investigators from the latter group tend to regard Type I and Type II errors as equally objectionable--hence, their more liberal criteria.

To illustrate some of the extant differences between "conservative criteria" and "liberal criteria," consider the ubiquitous liquid quantity conservative problem. We begin with two identical glasses containing equal amounts of water and then pour the contents of one glass into either a wider or narrower container. If this problem is administered according to the usual conservative criterion (e.g., Inhelder & Sinclair, 1969), subjects would be required to do at least three things: (a) judge whether or not the two quantities are still equal; (b) explain each judgment; (c) reply to a countersuggestion by the experimenter about each judgment. Any given administration of the problem is scored as a "pass" only if the subject judges that the quantities are still equal, provides an explanation which indicates that he grasps the underlying logic of the problem, and resists the experimenter's suggestion that the two quantities are in fact unequal. If the problem is administered according to the usual liberal criterion (e.g., Brainerd, 1974b), subjects would be required only to judge whether or not the two quantities are still equal after transformation. Any given administration of the problem is scored as a "pass" if the subject makes an equivalence judgment. The subject is said to possess conservation if the ratio of equivalence judgments to difference judgments departs significantly from chance expectations across a series of administrations of the problem.

The conservative criteria employed by orthodox investigators are open to the criticism that many subjects who possess the concepts being assessed will not be able to meet the stringent criteria. In the preceding illustration, for example, component _b_ is suspect. I have shown elsewhere on theoretical grounds that explanations introduce at least two specific sources of Type II error (Brainerd, 1973b, 1974a). Moreover, Siegel (1974a, 1974b) recently has shown that children's capacity to use the language tapped by component _b_ lags far behind their grasp of the concepts themselves. On the other hand, the more liberal criteria employed by less orthodox investigators are open to the criticism that some subjects who do not possess the concept being assessed will be able to satisfy the criteria. In the preceding illustration, for example, it is theoretically possible for a subject to adopt an equivalence response set and pass every trial without understanding conservation. More generally, Smedslund (1969) has argued that most Piagetian concept assessments are susceptible to "irrelevant hypotheses" (e.g., equivalence sets in the case of conservation) which allow subjects to generate correct judgments without possessing the concept. Although at present there does not appear to be any unequivocal empirical support for the existence and use of irrelevant hypotheses (cf. Brainerd, 1973b, Footnote 3, p. 176), many investigators have viewed Smedslund's argument as logically persuasive (e.g., Reese & Schack, 1974).

In short, researchers who study Piagetian concepts and seek to test predictions from the theory are confronted with the following dilemma: Is it better to adopt conservative Genevan criteria and run the risk of committing Type II errors or is it better to adopt more liberal criteria and run the risk of committing Type I errors? Does a criterion which may incorporate Type II error or a criterion which may incorporate Type I error provide a fairer test of the theory? To answer this question, we must know what the objective consequences of Type I and Type II errors are for the theory. We shall consider this problem below. First, the empirical significance of the criterion problem will be discussed as a means of narrowing its scope and sharpening its focus. We shall see that the criterion problem is of critical theoretical significance only in studies concerned with the order of emergence of two or more same-stage concepts. Next, the effects of Type I and Type II errors on expected frequencies for samples drawn from populations which do and do not support theoretical predictions will be reviewed. Contrary to the frequently acknowledged opinion that liberal criteria produce data which are less favorable to the Piagetian theory than conservative Genevan criteria (e.g., Gruen, 1966; Strauss, 1972), we shall see that Type I and Type II errors both favor the theory.

Empirical Significance of the Criterion Problem

In the cognitive-developmental literature, the criterion problem has been a source of controversy in the context of three specific questions: (a) the age of emergence of individual Piagetian concepts; (b) the trainability of individual Piagetian concepts; (c) the order of emergence of two or more same-stage Piagetian concepts. Concerning a, the age norms for concepts such as transitivity and conservation are two or three years older with Genevan criteria than with more liberal criteria (e.g., Braine, 1959; Braine & Shanks, 1965a, Brainerd, 1973a; Bryant & Trabasso, 1971). Concerning b, it has been suggested (Gruen, 1966; Kuhn, 1974; Strauss, 1972) that training effects observed with more liberal criteria may not be observed with Genevan criteria. Concerning c, two concepts belonging to the same Piagetian stage of mental development are less frequently observed to emerge in a fixed order with Genevan criteria than with more liberal criteria (cf. Brainerd, 1974a; Brainerd & Hooper, 1974).

Age norms, training effects, and emergence orders all have provoked their share of debate in the literature. However, disagreements in the literature notwithstanding, the first two questions are not especially crucial from the standpoint of the theory. Concerning age norms, it has been observed elsewhere that "age norms are trivial issues from the standpoint of Piagetian theory because the theory leaves wide latitude regarding them [Brainerd, 1973b, p. 173]." If, for example, it could be demonstrated that most concrete-operational concepts appear a full two years earlier than the nominal 7-to-8-year old norm, the theory would not be substantively affected. Second, concerning training effects, the suggestion that such effects may be observed with liberal criteria but not with conservative criteria would be important only if, as was once supposed (e.g., Mermelstein & Meyer, 1969; Smedslund, 1961), the theory specified that stage-related concepts cannot be trained. Although there has been some confusion in the literature over precisely what the predictions of the theory are vis-à-vis training, this issue has been clarified in recent years (Brainerd & Allen, 1971; Inhelder & Sinclair, 1969; Sinclair, 1973). Clearly, the theory predicts training effects (cf. especially Sinclair, 1973, pp. 57-58). Moreover, the conjecture that these effects are visible with liberal criteria but not with Genevan criteria has not received much support from neoPiagetian training experiments. In a recent review of this literature (Brainerd, 1973c), it was noted that, to date, choice of criterion has not been a critical variable in training experiments. The typical finding has been that a given treatment produces a training effect with <u>both</u> sorts of criteria, but the effect is more pronounced with liberal criteria.

Unlike a and b, the third question is crucial from the standpoint of the theory. Hence, let us be precise about what this question is concerned with. Three general types of asynchrony are possible with Piagetian concepts. First, there are between-stage asynchronies, which the theory terms vertical décalages. Between-stage asynchronies come in two varieties: within-concept and between-concept. Concerning the former, we are given two or more versions of the same general concept which ostensibly belong to different global stages of mental development and they are observed to emerge in a fixed order. Quantity conservation (concrete-operational stage) and density conservation (formal-operational stage) are a case in point. Concerning between-stage/between-concept asynchronies, we are given two or more distinct concepts which ostensibly belong to different stages and they are observed to emerge in a fixed order. Number conservation and implication reasoning are a case in point. With respect to both versions of vertical décalage, the prediction, of course, is that the item from the earlier stage invariably is acquired before the item from the later stage. The second type of asynchrony is within-stage/within-concept, which the theory terms horizontal décalage. For example, children are known to conserve in some nominally concrete-operational areas (e.g., number length) long before they conserve in others (e.g., weight, area). Although the theory makes no directional predictions about this type of asynchrony, it presently accomodates horizontal décalages on an ad hoc basis (cf. Flavell & Wohlwill, 1969). The third and most interesting form of asynchrony is within-stage/between-concept. Here, we are given two or more distinct concepts which ostensibly belong to the same stage and they are observed to emerge in a fixed order. The concrete-operational concepts of transitivity and conservation provide a standard illustration. It is with respect to within-stage/between-concept asynchrony that the criterion problem becomes of paramount importance.

Within-stage/between-concept asynchrony is more interesting than either between-stage or within-stage/within-concept asynchrony because it is generally proscribed by the theory (cf. Flavell & Wohlwill, 1969; Pinard & Laurendeau, 1969). Moreover, existing neoPiagetian models, which seek to refine Piaget's stages by adding quantifiable transition parameters (e.g., Case, 1972; Flavell & Wohlwill, 1969; Pascual-Leone, Pulos, & Parkinson, 1974; Pascual-Leone & Smith, 1969), also cannot accomodate this type of asynchrony. It is Piaget's structures-of-the-whole principle which precludes within-stage/between-concept asynchrony (Brainerd, 1974c; Pinard & Laurendeau, 1969). According to this principle, the concepts which characterize each global stage of mental development all presuppose the same set of tightly knit cognitive structures. During the course of each stage, the concepts identified with that stage are generated from the structures in much the same manner that a mathematician would deduce theorems from previously given proof procedures. Because the structures which define a given stage are so tightly knit, the theory stipulates that they must emerge synchronously. Within-stage/within-concept asynchronies are possible because the theory acknowledges the somewhat animistic possibility that certain content areas "resist" application of the structures more than other content areas. Thus, number conservation precedes quantity conservation because, for some unknown reason, quantity content resists application of the concrete-operational groupement structures more than number content does. However, when we consider two distinct concepts which are believed to presuppose the same underlying structures (i.e., belong to the same stage) within the same content area, the resistance argument cannot be invoked because content is held constant. If we compare transitivity of length with conservation of length, for example, the content area is the same and, hence, the two concepts should emerge synchronously (Brainerd, 1974c; Dagenais, 1973; Pinard & Laurendeau, 1969). If, instead, it is observed that one of the two concepts invariably precedes the other in most content areas, then the claim that both concepts presuppose the same level of cognitive structuration seems dubious.

The structures-of-the-whole principle notwithstanding, within-stage/between-concept asyncrhonies may indeed occur. They have been observed in profusion during recent years with concrete-operational concepts. Moreover, the asynchronies which have been reported seem to suggest a general underlying pattern which has been commented on elsewhere (Brainerd, 1974c; Burke-Merkle & Hooper, 1975). The structures which all concrete-operational concepts are said to presuppose are the so-called groupements. There are eight of these structures in all. Four of them are concerned with set-theoretic operations predicated on nested classes (cf. Piaget, 1949, pp. 109-125) and the remaining four are concerned with arithmetic operations predicated on asymmetrical and symmetrical relations (cf. Piaget, 1949, pp. 141-179). Hence, concrete-operational concepts tend to fall into two broad categories--those concerned primarily with reasoning about classes (e.g., class inclusion, double classification, cardinal number) and those concerned primarily with reasoning about relations (e.g., transitivity, double seriation, ordinal number). Suppose we think of the latter group of concepts as comprising a "relational dimension" and the former group of concepts as comprising a "classificatory dimension." Recent developmental evidence suggests that, contrary to structures-of-the-whole, children make considerable progress with the relational dimension before they make much progress with the classificatory dimension. Taken together, findings on four within-stage/between-concept asynchronies suggest this underlying pattern: transitivity vs. class inclusion (Brainerd, 1973a, 1974b; Brainerd & Vand Heuvel, 1974, double seriation vs. double classification (Hooper et al., 1974); ordinal number vs. cardinal number (Brainerd, 1973d, 1973e, 1974d; Brainerd & Fraser, 1975; Gonchar, 1974; Siegel, 1974c); relational groupement operations vs. classificatory groupement operations (Brainerd, 1972; bihoff, 1974; Weinreb & Brainerd, 1975). In each instance, the relevant relational concepts were observed to emerge before the relevant classificatory concepts.

Although the data reported in the studies just mentioned is both substantial and suggestive of an underlying pattern, liberal nonGenevan criteria were employed in all the studies. Therefore, it is possible for more orthodox investigators to invoke the Type I error criticism discussed earlier. Further doubts are raised by empirical evidence which suggests that some of the preceding asynchronies are not observed with conservative Genevan criteria (Dagenais, 1973; Gonchar, 1974; Lemerise, 1974; Sheppard, 1974; Smedslund, 1964). Moreover, Genevan criteria are known to reduce the possibility of finding asyncrhonies in same-stage concepts other than those mentioned above (Brainerd & Brainerd, 1972; Brainerd, 1974a; Brainerd & Hooper, 1974).

To summarize, we are confronted with two general problems. First, there is the debate over response criteria. Genevan-oriented investigators have argued that Type I errors are committed with liberal criteria and less orthodox investigators have argued that Type II errors are committed with conservative criteria. The second problem concerns the conflicting findings on within-stage/between-concept asynchrony. Recent evidence suggests that asynchronies of this theoretically crucial sort are observed more frequently with liberal criteria than with conservative criteria. Therefore, we cannot decide which findings are more believable or, vastly more important, how to proceed in future studies until we know precisely what effects Type I and Type II errors have on the null hypotheses tested in such studies. Is it possible, as Genevan reasoning implies, that Type I errors can manufacture asynchrony from synchrony? Is it possible, as the reasoning of other investigators implies, that Type II errors can mask real asynchrony? If one of these questions can be answered negatively and the other answered affirmatively, then we shall have a resolution of the criterion problem for studies concerned with within-stage/between-concept asynchrony.

Effects of Type I and Type II Errors on Different Populations

We must separately evaluate the effects of Type I and Type II errors on two different populations: theory-supportive popuations (i.e., those in which same-stage concepts emerge synchronously) and theory-contradictory populations (i.e., those in which same-stage concepts emerge asynchronously). With theory-supportive populations, we are especially concerned to know whether or not Type I error increases the probability of obtaining asynchronous data. With theory-contradictory populations, we are especially concerned to know whether or not Type II error increases the probability of obtaining synchronous data.

The structure of a concept development study may be described as follows. We are given two (or more) concepts, A and B, which are known to emerge during some age range R. Tests of A and B are administered to a sample from R. The subjects' responses are scored according to a a conservative or liberal criterion and the scores are used to assign each subject to one of four categories: pass A-pass B (A/B), pass A-fail B (A/~B), fail A-pass B (~A/B), and fail A-fail B (~A/~B). The appropriate null hypothesis is that the observed frequencies of the second and third categories do not differ significantly. An exact significance test for this hypothesis is given by

$$P(\underline{x}) = \binom{N}{\underline{x}} P^{\underline{x}} Q^{N-\underline{x}}, \qquad [1]$$

where $\underline{x}$ = either the number of A/~B subjects or the number of ~A/B subjects, N = the total number of A/~B subjects and ~A/B subjects, P = .50, and Q = .50. For three or more concepts, Eq. 1 provides an exact significance test for each of the pairwise null hypotheses. Approximate tests of significance for three or more concepts are provided by less cumbersome scalometric procedures (e.g., Green, 1954). The important fact to bear in mind is that subjects classified as A/B or ~A/~B are irrelevant to the null hypothesis and, hence, we must concern ourselves with the effects of Type I and Type II criterion errors on the relative observed frequencies of categories A/~B and ~A/B.

<u>Theory-Supportive Populations: Null Hypothesis True</u>

Assume that the tests are administered to a sample drawn from a population in which A and B emerge synchronously. The population frequencies for the four possible types of subjects are: P(A/B) = P_1, P(A/~B) = P_2, P(~A/B) = P_2, and P(~A/~B) = P_3, where $P_1 + 2P_2 + P_3 = 1$. Because A and B are not acquired in any particular order, it must be the case that their population frequencies are equal. Suppose that the responses of our subjects are scored according to a liberal criterion with a Type I error rate of $0 < \underline{x} < 1$. How will this affect the expected frequencies of the four subject categories?

Concerning category A/B, the probability that any subject who belongs to this category will be correctly classified by our criterion is unity. For category A/~B, the probability that any subject who belongs to this category will be correctly classified by our criterion is 1 - x. The probability that an A/~B subject will be incorrectly assigned to the A/B category is x. For category ~A/B, the probability that any subject who belongs to this category will be correctly classified by our criterion is 1 - x. The probability that a ~A/B subject will be incorrectly assigned to the A/B category is x. For category ~A/~B, the probability that any subject who belongs to this category will be correctly classified by our criterion is $1 - 3x^2$. The probability that a ~A/~B subject will be incorrectly assigned to categories A/B, A/~B, and ~A/B, respectively, is x^2 in each case. Hence, the expected frequencies of the four categories for any sample drawn from this population are :

$$E(A/B) = P_1 + 2xP_2 + x^2P_3, \qquad [2]$$

$$E(A/\sim B) = P_2[1 - x] + x^2P_3, \qquad [2]$$

$$E(\sim A/B) = P_2[1 - x] + x^2P_3, \text{ and} \qquad [4]$$

$$E(\sim A/\sim B) = P_3[1 - 3x^2]. \qquad [5]$$

Note that if $P_1 + 2P_2 + P_3 = 1$, then Eq. 2-Eq. 5 must sum to 1 also.

The principal effect of Type I errors with theory-supportive samples is on the first and last categories. The expected value for the first category is substantially increased and the expected value for the last category is substantially decreased, relative to their population values. As mentioned earlier, however, we are concerned primarily with the effects of criteria errors on the second and third categories. Note that, from the standpoint of the null hypothesis tested in concept development studies, Type I errors are of no consequence. Although Type I errors decrease the expected values of both A/~B and ~A/B relative to their population values, both are decreased by the same amount. Therefore, if the population frequencies of the two categories are equal, as they must be if the population is theory-supportive, then their expected frequencies after Type I error also are equal. Because the null hypothesis is concerned with the <u>difference</u> between the observed freqencies for the two categories, Type I criterion errors neither increase nor decrease the probability of obtaining asynchronous data given synchrony in the population.

Now, suppose that our subjects' responses are scored according to a conservative criterion with a Type II error rate of $0 < y < 1$. The effect is essentially the reverse of the above. Again, assume that the population frequencies of the four categories are P_1, P_2, P_2, and P_3, respectively. For category A/B, the probability that any subject who belongs to this category will be correctly classified by our criterion is $1 - 3y^2$. The probability that an A/B subject will be incorrectly assigned to categories A/~B, ~A/B, and ~A/~B is y^2 in each case. For category A/~B, the probability that any subject who belongs to this category will be correctly classified by our criterion is $1 - y$. The probability that an A/~B subject will be incorrectly assigned to the ~A/~B category is y. For category ~A/B, the probability that any subject who belongs to this category will be correctly classified by our criterion is $1 - y$. The probability that a ~A/B subject will be inocrrectly assigned to the ~A/~B category is y. Finally, for category ~A/~B, the probability that any subject who belongs to this category will be correctly classified by our criterion is unity. Hence, the expected frequencies for the four categories are

$$E(A/B) = P_1[1 - 3y^2], \qquad [6]$$

$$E(A/{\sim}B) = P_2[1 - y] + y^2P_1, \qquad [7]$$

$$E({\sim}A/B) = P_2[1 - y] + y^2\,P_1, \text{ and} \qquad [8]$$

$$E({\sim}A/{\sim}B) = P_3 + 2yP_2 + y^2P_1. \qquad [9]$$

As was the case for Type I errors, the principal effect of Type II errors, given a theory-supportive sample, is to inflate the ~A/~B category substantially and deflate the A/B category substantially. As was also the case for Type I errors, Type II errors do not differentially affect the critical A/~B and ~A/B categories. If the frequencies of these two categories are equal in the population, then their expected frequencies after Type II error also must be equal. Therefore, Type II errors neither increase nor decrease the probability of obtaining asynchronous data given synchrony in the population.

Theory-Contradictory Population: Null Hypothesis False

Assume that the tests are administered to a sample from a population in which A and B emerge in a fixed order and assume that the order is A > B. The population frequencies for the four possible types of subjects are: $P(A/B) = P_1$, $P(A/{\sim}B) = P_2$, $P({\sim}A/B) = 0$, and $P({\sim}A/{\sim}B) = P_3$, where $P_1 + P_2 + P_3 = 1$. Suppose that the responses of our subjects are scored according to the same literal criterion as before. The probability that any A/B subject will be correctly classified is unity. The probability that any A/~B subject will be correctly classified is $1 - x$ and the probability that he will be incorrectly classified as an A/B is x. The probability that any ~A/~B subject will be correctly classified is $1 - 3x^2$ and the probability that he will be incorrectly classified as an A/B or an A/~B or a ~A/B is x^2 in each case. Hence, the expected frequencies for the four categories are:

$$E(A/B) = P_1 + \underline{x}P_2 + \underline{x}^2P_3, \quad [10]$$

$$E(A/\bar{}B) = P_2[1 - \underline{x}] + \underline{x}^2P_3, \quad [11]$$

$$E(\bar{}A/B) = \underline{x}^2P_3, \text{ and} \quad [12]$$

$$E(\bar{}A/\bar{}B) = P_3[1 - 3\underline{x}^2]. \quad [13]$$

Unlike theory-supportive samples, Type I errors do have a differential effect on the crucial A/~B and ~A/B categories with theory-contradictory samples. Explicitly, Type I errors tend to increase the observed frequency of ~A/B relative to the observed frequency of A/~B: Given that P_2, P_3, and $\underline{x}$ all lie between 0 and 1, the difference between the population frequencies of the two categories (P_2 and 0) must be greater than the difference between the expected frequencies of the two categories after Type I error ($P_2[1 - x] + \underline{x}^2P_3$ and $\underline{x}^2P_3$). Therefore, the commission of Type I errors will <u>increase</u> the probability of falsely accepting the null hypothesis that A/~B and ~A/B subjects occur with equal frequency in the population. Given a constant error rate, this masking effect will become more serious as the absolute magnitude of P_2 decreases. If A → B is a "robust" sequence (i.e., several years elapse between the onset of A and the first evidence of B), then the frequency of A/~B subjects in the population will be substantial relative to the frequencies of A/B and ~A/~B subjects. If A → B is a "precise" sequence (i.e., only a few months elapse between the onset of A and the first evidence of B), then the frequency of A/~B subjects in the population will be small relative to the frequencies of A/B and ~A/~B subjects. Hence, the consequences of making Type I errors will be more serious for precise sequences such as identity vs. equivalence (cf. Brainerd & Hooper, 1974) and multiple classification vs. multiple seriation (cf.

Brainerd, 1974c) than for robust sequences such as transitivity vs. class inclusion (cf. Brainerd, 1973a).

Now, suppose that we employ the conservative criterion mentioned earlier. The population frequencies of the four categories again are assumed to be P_1, P_2, 0, and P_3, respectively. The probability that any A/B subject will be correctly classified is $1 - 3y^2$ and the probability that he will be incorrectly classified as an A/~B or a ~A/B or a ~A/~B is y^2 in each instance. The probability that any A/~B subject will be correctly classified is $1 - y$ and the probability that he will be incorrectly classified as a ~A/~B is y. The probability that any ~A/~B subject will be correctly classified is unity. Hence, the expected frequencies for the four categories are:

$$E(A/B) = P_1[1 - 3y^2], \qquad [14]$$

$$E(A/\sim B) = P_2[1 - y] + y^2 P_1, \qquad [15]$$

$$E(\sim A/B) = y^2 P_1, \text{ and} \qquad [16]$$

$$E(\sim A/\sim B) = P_3 + yP_2 + y^2 P_1. \qquad [17]$$

Thus, the effect of Type II errors on theory-contradictory samples is precisely the same as the effect of Type I errors. Type II errors increase the observed frequency of ~A/B subjects relative to the observed frequency of A/~B subjects and, therefore, the probability of falsely accepting the null hypothesis that A/~B and ~A/B subjects occur with equal frequency in the population also increases. As was also the case for Type I errors, the practial consequences of committing Type II errors increase as P_2 decreases. Given a constant Type II error rate, its consequences will be more serious for precise sequences than for robust ones.

Conclusions and Recommendations

The preceding analysis has several interesting implications for the criterion problem with which we began. First, the claim that liberal criteria which incorporate Type I error tend to produce findings which are overly unfavorable to the theory obviously is false. Given that a population is theory-supportive (null hypothesis true), Type I errors neither increase nor decrease the probability of incorrectly rejecting the null hypothesis. Moreover, given that a population is theory-contradictory (null hypothesis false), Type I errors actually reduce the probability of correctly rejecting the null hypothesis. In short, regardless of whether or not liberal criteria do in fact incorporate Type I errors, the use of such criteria cannot possibly militate unfairly against Piagetian predictions. Quite to the contrary, if it happens that Type I errors are committed with a certain criterion, then the chances of finding spurious support for the theory are enhanced. A second implication of the preceding analysis is that the claim that conservative criteria tend to produce findings which unfairly favor the theory obviously is true. Given that a population is theory-contradictory, Type II errors increase the probability that the null hypothesis will be incorrectly accepted. As is the case for Type I errors, therefore, Type II errors increase the chances of obtaining spurious support for the theory.

There is one very positive outcome of the preceding analysis which should not go unmentioned. The analysis indicates that there is no psychometric foundation for the somewhat nihilistic view that obtaining or not obtaining theory-supportive findings in concept development studies is largely a function of arbitrary decisions about response criteria. This view, which has been expressed with distressing regularity in recent years (e.g., Hooper, 1974; Kuhn, 1974), undoubtedly has been a source of considerable uncertainty to concept development researchers. A rapprochement now may be effected. Psychometrically speaking, it is impossible to manufacture spuriously asynchronous data via the commission of either Type I or Type II criterion errors. Hence, evidence of asynchrony which presently exists in the literature cannot be dismissed on the grounds that the response criteria employed in the relevant studies incorporated either Type I or Type II errors, although it is always possible that this evidence could be dismissed on some other grounds. On the other hand, it is possible to manufacture spuriously synchronous data via the commission of Type I and Type II errors. Therefore, we must be more circumspect about accepting evidence of synchrony than we are about accepting evidence of asynchrony.

An important and provocative implication of what has just been said is that, from the standpoint of the criterion debate, existing studies in which asynchronies have been observed are intrinsically more believable than existing studies in which synchronies have been observed. Earlier, it was noted that substantial evidence of theory-contradictory asynchrony exists for concrete operational concepts such as transitivity, class inclusion, multiple seriation, multiple classification, ordinal number, cardinal number, and groupement operations using liberal criteria. It also was noted that some of these same asynchronies have not been observed in studies where conservative criteria were employed. We now know that Type I criterion errors cannot possibly explain findings of the former sort. If Type I errors were made in these studies, then the reported asynchronies appeared in spite of them rather than because of them. We also know that Type II errors can explain findings of the latter sort--at least in part. In view of the fact that Type I and Type II errors both tend to produce spuriously synchronous data, it is psychometrically reasonable to suppose that the effective Type I error rates of the liberal criteria employed in the first group of studies must be substantially nearer to zero than the effective Type II error rates of the conservative criteria employed in the second group of studies.

References

Beilin, H. The training and acquisition of logical operations. In M. F. Rosskopf, L. P. Steffe, & S. Taback (Eds.), Piagetian cognitive developmental research and mathematical education. Washington: National Council of Teachers of Mathematics, 1971.

Braine, M. D. S. The ontogeny of certain logical operations: Piaget's fomulation examined by nonverbal methods. Psychological Monographs, 1959, 73 (Whole No. 475).

Braine, M. D. S. Piaget on reasoning: A methodological critique and alternative proposals. In W. Kessen & C. Kuhlman (Eds.), Thought in the young child. Monographs of the Society for Research in Child Development, 1962, 27 (Whole No. 83).

Braine, M. D. S. Development of a grasp of transitivity of length: A reply to Smedslund. Child Development, 1964, 35, 799-810.

Braine, M. D. S. & Shanks, B. L. The conservation of a shape property and a proposal about the origin of the conservations. Canadian Journal of Psychology, 1965, 19, 197-207. (a)

Braine, M. D. S. & Shanks, B. L. The development of conservation of size. Journal of Verbal Learning and Verbal Behavior, 1965, 2, 227-242. (b)

Brainerd, C. J. Structures of thought in middle-childhood: Recent research on Piaget's concrete-operational groupements. Paper read at the Third Annual Conference on Structural Learning, Philadelphia, March-April, 1972.

Brainerd, C. J. Order of acquisition of transitivity, conservation, and class inclusion of length and weight. Developmental Psychology, 1973, 8, 105-116 (a)

Brainerd, C. J. Judgments and explanations as criteria for the presence of cognitive structures. Psychologica' Bulletin, 1973. 79, 172-179. (b)

Brainerd, C. J. NeoPiagetian training experiments revisited: Is there any support for the cognitive-developmental stage hypothesis? Cognition, 1973, 2, in press. (c)

Brainerd, C. J. Mathematical and behavioral foundations of number. Journal of General Psychology, 1973, 88, 221-281, (d)

Brainerd, C. J. The origins of number concepts. Scientific American, 228 (3), 101-109 (e)

Brainerd, C. J. Postmortem on judgments, explanations, and Piagetian cognitive structures. Psychological Bulletin, 1974, 81, 70-71. (a)

Brainerd, C. J. Training and transfer of transitivity, conservation and class inclusion of length. Child Development, 1974, 45, 324-334. (b)

Brainerd, C. J. Structures-of-the-whole: Is there any glue to hold the concrete-operational "stage" together? Paper read at Canadian Psychological Association, Windsor, June, 1974. (c)

Brainerd, C. J. & Allen, T. W. Experimental inductions of the conservation of "first-order" quantitative invariants. Psychological Bulletin, 1971, 75, 128-144.

Brainerd, C. J., & Brainerd, S. H. Order of acquisition of number and liquid quantity conservation. Child Development, 1972, 43, 1401-1406.

Brainerd, C. J., & Hooper F. H. A methodological analysis of developmental studies of identity and equivalence conservation. Unpublished manuscript University of Alberta, 1974.

Brainerd, C. J., & Vanden Heuvel, K. Development of geometric imagery in five-to-eight-year-olds. Genetic Psychology Monographs, 1974, 89, 98-143.

Bruner, J. S. The course of cognitive growth. American Psychologist, 1964, 19, 1-15.

Bryant, P. E., Reply to Youniss and Furth. Nature, 1973, 244, 315-316.

Bryant, P. E., & Trabasso, T. Transitive inferences and memory in young children. Nature, 1971, 232, 456-458.

Burke-Merkle, A., & Hooper, F. H. Logical task performances in the pre-school: The effects of classification and seriation instruction. American Educational Research Journal, 1975, in press.

Case, R. Validation of a neo-Piagetian mental capacity construct. Journal of Experimental Child Psychology, 1972, 14, 287-302.

Dagenais, Y. Analyse de la cohérence opératoire entre les groupements d'addition des classes, de multiplication des classes et d'addition des relations asymétriques. Unpublished doctoral dissertation Université de Montréal, 1973.

Dihoff, R. E. Multidimensional scaling of Piagetian data. Unpublished master's thesis, University of Wisconsin, 1974.

Flavell, J. H., & Wohlwill, J. F. Formal and functional aspects of cognitive development. In D. Elkind & J. H. Flavell (Eds.), Studies in cognitive development. New York: Oxford, 1969.

Gonchar, A. The development of number concepts. Unpublished master's thesis, University of Wisconsin, 1974.

Green, B. F. Attitude measurement. In G. Lindzey (Ed.), Handbook of social psychology. Vol. 1. Cambridge, Mass.: Addison-Wesley, 1954.

Gruen, G. E. Note on conservation: Methodological and definitional considerations. Child Development, 1966, 37, 977-983.

ERIC

Hooper, F. H. Discussion. Quantity and logical concept development: Alternatives and extensions of the Piagetian model. Symposium presented at Canadian Psychological Association, Windsor, Ontario, June, 1974.

Hooper, F. H., Sipple, T. S., Goldman, J. A., & Swinton, S. S. A cross-sectional investigation of children's classificatory development. Technical Report No. 295, Research and Development Center for Cognitive Learning, University of Wisconsin, 1974.

Inhelder, B., Bovet, M., Sinclar, H., & Smock, C. D. On cognitive development. American Psychologist, 1966, 21, 1-23.

Inhelder, B., & Sinclair, H. Learning cognitive structures. In P. H. Mussen, J. Langer, & M. Covington (Eds.), Trends and issues in developmental psychology. New York: Holt, Rinehart, & Winston, 1969.

Kuhn, D. Inducing development experimentally: Comments on a research paradigm. Developmental Psychology, 1974, 11, in press.

Lasry, J. C. & Laurendeau, M. Apprentissage empirique de la notion d'inclusion. Human Development, 1969, 12, 141-153.

Lemerise, T. Etude experimentale du modele Piagetian de l'origine du nombre par une methode d'apprentissage provoque. Paper read at Canadian Psychological Association, Windsor, Ontario, June, 1974.

Mehler, J., Bever, T. G. Cognitive capacity of very young children. Science, 1967, 158, 141-142.

Mermelstein, E., & Meyer, E. Conservation training techniques and their effects on different populations. Child Development, 1969, 40, 471-490.

Pascual-Leone, J., Pulos, S. M., & Parkinson, G. M. Constructive abstraction (structural learning) and concept development. Unpublished manuscript, York University, 1974.

Pascual-Leone, J., & Smith, J. The encoding the decoding symbols by children. Journal of Experimental Child Psychology, 1969, 8, 3238-355.

Piaget, J. Traité de logique. Paris: Colin, 1949.

Pinard, A., & Laurendeau, M. "Stage" in Piaget's cognitive developmental theory: Exegesis of a concept. In D. Elkind & J. H. Flavell (Eds.), Studies in cognitive development. New York: Oxford, 1969.

Reese, H. W., & Schack, M. L. Comment on Brainerd's criteria for cognitive structures. Psychological Bulletin, 1974, 81, 67-69.

Riley, C. A., & Trabasso, T. Comparatives, logical structures, and encoding in a transitive inference task. Journal of Experimental Child Psychology, 1974, 17, 187-203.

Sheppard, J. L. Concrete operational thought and developmental aspects of solutions to a task based on a mathematical three group. Developmental Psychology, 1974, 10, 116-123.

Siegel, L. S. The relationship of language and thought in the young child: The case of quantity concepts. Unpublished manuscript, MacMaster University, 1974. (a)

Siegel, L. S. The relationship of language and thought in the preoperational child: A reconsideration of nonverbal alternatives to Piagetian tasks. Paper read at Canadian Psychological Association, Windsor, Ontario, June, 1974, (b)

Siegel, L. S. Development of number concepts: Ordering and correspondence operations and the role of length cues. Developmental Psychology, 1974, 11, in press. (c)

Sinclair, H. Recent Piagetian research in learning studies. In M. Schwebel & J. Raph (Eds.), Piaget in the classroom. New York: Basic Books, 1973.

Smedslund, J. The acquisition of conservation of substance and weight in children. I. Introduction. Scandanavian Journal of Psychology, 1961, 2, 11-20.

Smedslund, J. Development of concrete transitivity of length in children. Child Development, 1963, 34, 389-405.

Smedslund, J. The development of transitivity of length: A comment on Braine's reply. Child Development, 1965, 36, 577-580.

Smedslund, J. Psychological diagnostics. Psychological Bulletin, 1969, 71, 234-248.

Strauss, S. Inducing cognitive development and learning: A review of short-term training experiments. I. The organismic-developmental approach. Cognition, 1, 329-357.

Weinreb, N., & Brainerd, C. J. A developmental study of Piaget's groupement model of the emergence of speed and time concepts. Child Development, 1975, 46, in press.

Youniss, J., & Furth, H. G. Reasoning and Piaget: A comment on Bryant and Trabasso. Nature, 1973, 244, 314-315.