

## The Logic of Interpreting Evidence of Developmental Ordering: Strong Inference and Categorical Measures

James A. Dixon  
College of William and Mary

Colleen F. Moore  
University of Wisconsin—Madison

Developmental ordering is a fundamental prediction of developmental theories and a central issue in developmental research. However, logically sound evidence of developmental ordering is difficult to obtain. This article analyzes the logical basis of testing developmental order hypotheses with categorical measures. Depending on whether saltatory (i.e., discrete) or continuous developmental changes are being assessed, the observed relationship between categorical measures yields very different types of information about developmental ordering. When change is continuous, the relationship between the measures does not confirm any one ordering hypothesis, but rather, disconfirms one or more hypotheses. Whether an underlying variable undergoes saltatory or continuous development has long been recognized as an important theoretical issue, but its impact on the interpretation of developmental ordering has not previously been explicated.

Developmental theories live and die by the orderings they predict. The order in which skills, abilities, or knowledge structures emerge is perhaps the most important type of evidence in developmental psychology. The classic structuralist approaches to development, such as Piaget's, obviously rely heavily on developmental order, but the concept of developmental ordering is fundamental to most developmental research and theories, structuralist or not.

Consider a few prominent examples of developmental ordering hypotheses from various areas. Simon and Klahr's (1995) computational model of the development of number conservation predicts that the ability to measure must be developmentally prior to conservation. Similarly, Karmiloff-Smith's (1991) representational-redescription hypothesis predicts that mastery of a skill must precede its representational redescription. Case and Okamoto's (1996) theory of central conceptual structures predicts developmental synchrony among items that use the same conceptual structure. Likewise, Perner (1988) argued that the ability to simultaneously consider and coordinate two representations must precede understanding of false belief, an important index of theory of mind. Lewis, Sullivan, Stranger, and Weiss (1989) hypothesized that the emotion of fear would develop prior to self-recognition but that embarrassment would develop either at the same time as or after self-recognition. Xu and Carey (1996) hypothesized that infants first construct the concept for "object" on the basis of spatiotemporal information and later construct concepts for indi-

vidual objects on the basis of their particular properties. Despite the considerable diversity in the content of these theories and hypotheses, they all face the same analytical problem: how to test predictions about developmental ordering.<sup>1</sup>

Unfortunately, the most obvious test of developmental ordering, directly comparing means or scores on two measures, does not yield much information about developmental order. The problem is that researchers rarely know if two measures of different constructs are comparable (Chapman & Chapman, 1973, 1978). For example, Newcombe, Huttenlocher, Drummey, and Wiley (1998) assessed children's ability to use two forms of coding spatial location: dead reckoning and place learning. With dead reckoning, spatial location is coded using optic-flow, vestibular, and kinesthetic information about one's position in 3-D space. With place learning, spatial location is coded with reference to landmarks. Children were asked to retrieve an object that they had just watched the experimenter bury in a sandbox. Newcombe et al. found that when landmarks were unavailable, and therefore dead reckoning was required, the performance of 28–36-month-old children was quite poor (although significantly better than chance). When landmarks were available, and therefore place learning could be used, performance was significantly better.<sup>2</sup> At first glance, this result may seem to imply that place learning is more developmentally advanced than dead reckoning. However, as Newcombe et al. noted, dead reckoning is affected by the degree of rotation and translation in space. Place learning is affected by the availability and salience of external landmarks. Unless these factors were adjusted so that the task was equally difficult using either process, comparing the absolute level of accuracy across the

---

James A. Dixon, Department of Psychology, College of William and Mary; Colleen F. Moore, Department of Psychology, University of Wisconsin—Madison.

Preparation of this manuscript was supported in part by Grant SBR-9874648 from the National Science Foundation.

We thank Nazan Aksan, Ashley Bangert, Charles J. Brainerd, and Adam Rubenstein for helpful comments on a previous version of the article.

Correspondence concerning this article should be addressed to James A. Dixon, Department of Psychology, College of William and Mary, P.O. Box 8795, Williamsburg, Virginia 23187-8795. Electronic mail may be sent to [jadixo@wm.edu](mailto:jadixo@wm.edu).

---

<sup>1</sup> Note that the examples presented here are empirically falsifiable sequences rather than "measurement sequences." In measurement sequences, the predicted developmental ordering is a logical consequence of the later developing item's containing the prior developing item. (See Brainerd, 1993.)

<sup>2</sup> Newcombe et al. (1998) also tested a sample of younger children aged 16 to 24 months. The major focus of their analysis was on the change in the use of place learning within this younger group.

		Skill B	
		N	Y
Skill A	Y	O	X
	N	X	O

**Synchrony**

		Skill B	
		N	Y
Skill A	Y	X	X
	N	X	O

**Priority**

		Skill B	
		N	Y
Skill A	Y	O	X
	N	X	X

**Priority**

*Figure 1.* Each contingency table shows a pattern usually interpreted as evidence of developmental ordering. People are classified as either having (Y = yes) or not having (N = no) each of two skills, A and B. The X's refer to the presence of people in the cell. The O's refer to the absence of people in the cell. The upper panel shows the pattern usually interpreted as developmental synchrony. The lower left panel shows the pattern usually interpreted as developmental priority of Skill A and Skill B. The lower right panel shows the pattern usually interpreted as developmental priority of Skill B over Skill A.

two conditions cannot provide information about developmental order. The more general point, of course, is that without extensive standardizing of measures to ensure comparability, directly comparing the means or scores of two measures will not provide adequate tests of developmental ordering (Chapman & Chapman, 1973, 1978).

Researchers have also tested developmental ordering by examining the distribution of people's scores on the two measures across development (Flavell, 1971; Guttman, 1944; Lewis et al., 1989; Wohlwill, 1973). In this approach, individuals across the developmental span of interest are sampled, and two measurements are taken at the same time for each individual. The relationship between the two measures across the developmental period of interest is used to draw conclusions about developmental ordering. The lower left panel of Figure 1 gives a very simple example of this type of evidence in a contingency table. Assume that Skill A and Skill B have been measured in the same session of an experiment and that the sample of children spans the age range across which these skills develop. For example, in studying emotion and self-recognition, Lewis et al. (1989) sampled infants ranging in age from 9 to 24 months and administered measures of self-recognition and emotion in the same session. The label Y (yes) in Figure 1 indicates that the skill is present; N (no) indicates that it is absent. The X's in the figure denote the presence of people in a cell. The O's denote the absence of people. (The O's can also be considered to denote the presence of a small number of people attributable to error; however, for simplicity of presentation, we refer to O's as the absence of people throughout this article.) This pattern is usually interpreted as evidence supporting the hypothesis

that Skill A develops before Skill B. The reasoning is straightforward. Some individuals have been sampled early in development before they acquire either skill. These people appear in the lower left cell. Other people are caught later in development. These people have Skill A but not Skill B (upper left cell). The upper right cell shows people who are at the highest level of development; they have both skills. Importantly, there is an absence of people who have Skill B but do not have Skill A (lower right cell).

Inferences about developmental ordering can be made with categorical measures, as in the example above, or with continuous measures (e.g., measures that assess the extent to which Skill A or Skill B has developed). Dixon (1998) showed that when researchers use continuous measures, the observed relationship between the two measures does not necessarily reflect the relationship between the underlying variables.<sup>3</sup> However, the observed relationship between measures can effectively disconfirm one of the two possible priority hypotheses. The extent to which the observed relationship between scores eliminates developmental ordering hypotheses depends on whether the measures are ordinal or interval scales, a topic to which we return later in the article. (See Dixon, 1998, for details.)

The observed relationship between categorical measures can be interpreted much more directly but only if the development of the underlying variables is discrete (i.e., abruptly shifts from one state

<sup>3</sup> We are using the term *underlying variable* to refer to the psychological entity or construct being measured. We distinguish between the underlying variable and the measurement of it, as do virtually all psychological theories.

to another). Categorical measures do not, of course, necessarily imply that the underlying variables themselves change categorically. If a continuously developing variable is assessed with a categorical measure, the categorical measure essentially imposes a success–failure cutoff point on the underlying variable (see Brainerd, 1977, and Brainerd & Hooper, 1975, for a discussion of issues in placing the cutoff point). For example, Cornell, Heth, and Alberts (1994), in a study of place recognition, asked children who were either on or off a route they had just traveled, whether or not they were on their original path. As Cornell et al.'s analysis showed, children's impression of being on or off route was a continuous dimension that depended on the familiarity of their surroundings. Their categorical answers (i.e., yes or no) reflected the categorical nature of the measure rather than two discrete states of mind (i.e., "I'm on route" or "I'm off route"). The point here is simply that the use of categorical measures does not imply that the underlying variable is categorical; variables that undergo continuous change and development can also be assessed with categorical measures.<sup>4</sup>

Below we analyze the profound logical implications for interpreting evidence of developmental order when categorical measures of continuously developing variables are used. First, we briefly outline the assumptions necessary to interpret the relationship between categorical measures. These are the same basic assumptions researchers implicitly accept in order to compare means or to make any other comparison of measures. Second, we describe how categorical measures of variables that undergo discrete developmental change can be used to infer developmental order. Third, we describe how using categorical measures of variables that undergo continuous developmental change has surprising implications for inferring developmental order and requires careful application of disconfirmatory logic. Finally, we compare the results of this analysis of categorical measures with that of our previous analysis of continuous measures (Dixon, 1998) and show that categorical and continuous measures yield similar types of evidence with regard to developmental ordering unless a set of stringent conditions is met.

#### Basic Assumptions

To simplify the presentation, assume that the people across the span of development have been measured in a cross-sectional design (longitudinal designs do not change the conclusions, as we discuss later). Each individual is measured once on each skill, and those measurements are taken at the same time or as close in time as is reasonably possible. In order to compare the scores on those two measures, researchers must first assume that the measures have acceptable validity in the sense that they capture change in the underlying variable of interest and change only in that variable. (Any comparisons of measures are problematic without this assumption of validity; see Campbell & Stanley, 1963; Cook & Campbell, 1979; Krathwohl, 1985.) Second, researchers interpret an observed relationship only if it is statistically significant. Methods for testing the statistical significance of relationships between categorical variables are readily available (e.g., Hildebrand, Laing, & Rosenthal, 1977; see Wickens, 1998, for a recent review).

#### Categorical Measures of Discrete Developmental Change

Figure 1 shows three patterns in a contingency table. These patterns are usually interpreted as evidence of developmental or-

der. The upper panel in Figure 1 shows the developmental pattern that is usually interpreted as developmental synchrony. The explanation for this data pattern is that some people have been sampled before they develop either skill (the lower left cell), and other people, after they have developed both skills (the upper right cell). Importantly, people who have one skill but not the other are not observed, which is consistent with the hypothesis that the skills develop synchronously.

The lower left panel of Figure 1 shows the developmental pattern usually interpreted as the developmental priority of Skill A over Skill B. As described above, the explanation for this pattern is that we observe people in three developmental states: Those at the least advanced level have not developed either skill, those at the next most advanced level have developed Skill A but not Skill B, and people at the most advanced level have developed both skills. People who have Skill B but not Skill A are not observed. The lower right panel shows the reverse pattern, which is usually interpreted as Skill B developing prior to Skill A. The explanation is analogous to the one for the lower left panel.

When the skills change discretely from absent to present, the relationship between the two categorical measures in a contingency table has a straightforward interpretation. The standard interpretation of each pattern discussed above is appropriate. Each pattern is consistent with one developmental relationship and disconfirms the other two relationships.

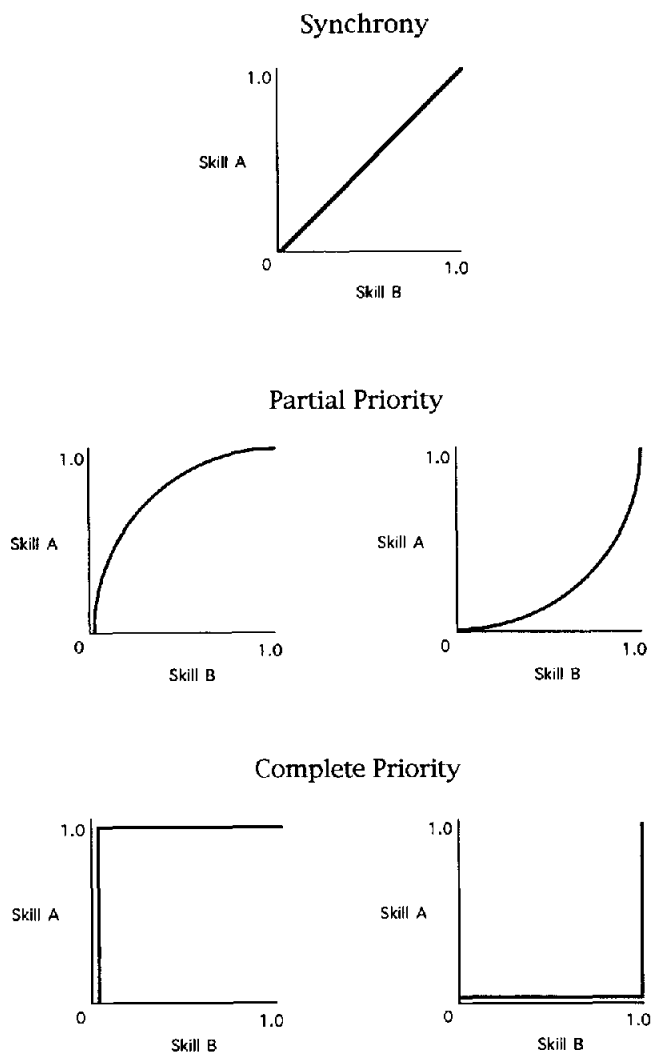
#### Categorical Measures of Continuous Developmental Change

When underlying variables are believed to undergo continuous developmental change, the developmental ordering hypotheses must be defined more specifically, because developmental ordering can have a variety of meanings (Flavell, 1971; Wohlwill, 1973). For example, one might define developmental priority as (a) one skill starting development before a second skill starts to develop or (b) one skill completing development before another skill starts development.

Figure 2 presents five idealized developmental patterns. Each pattern graphically represents a different developmental ordering hypothesis. The top panel shows one definition of developmental synchrony. Both skills begin development at the same time and develop at the same rate. The panels in the middle row show two partial developmental priority hypotheses. One skill undergoes considerable development before the other skill begins to develop. The panels in the bottom row show two complete priority hypotheses. One skill both starts and completes development before the other skill starts to develop.

Obviously, this is not an exhaustive set of developmental ordering hypotheses, but these hypotheses are often proposed in developmental research. For example, Kail (1991, 1997) proposed that a single global factor is responsible for changes in speed of processing. Performance on a wide range of tasks is predicted to show developmental synchrony because growth of a single factor is responsible for the developmental changes. Reyna and Brainerd (1995) proposed a developmental shift from verbatim-based to gist-based reasoning. Early in development, children predomi-

<sup>4</sup> The converse, of course, is also true; continuous measures do not imply continuous underlying variables.



*Figure 2.* Each idealized pattern shows a hypothesized underlying developmental ordering between two skills, A and B. In each panel, development of Skill A is represented on the vertical axis and development of Skill B, on the horizontal axis. The top panel shows a synchrony hypothesis: Skills A and B develop simultaneously and at the same rate. The middle panels show two partial priority hypotheses: Skill A develops more quickly at first than Skill B (left side), and Skill B develops more quickly at first than Skill A (right side). The bottom panels show two complete priority hypotheses: On the left side, Skill A both starts and completes development before Skill B begins development; on the right side, Skill B starts and completes development before Skill A begins development.

nantly use verbatim-based reasoning, although they have some ability to use gist-based reasoning as well. This hypothesis specifies partial developmental priority: Verbatim-based reasoning develops more quickly than gist-based reasoning. Gist-based reasoning eventually reaches mature levels (and becomes the preferred method of reasoning). Karmiloff-Smith (1991) proposed that children must achieve behavioral mastery of a skill (e.g., grammatical use of language) before that knowledge can be redescribed and made available as "data" to other parts of the mind (e.g., explicit knowledge of grammatical rules). This hypothesis clearly specifies complete developmental priority: Behavioral competence must

complete development before explicit knowledge of the skill can develop.

Portraying the developmental ordering hypotheses as relationships between two continuous underlying variables may seem to suggest that the use of categorical measures is simply wrong-headed, especially given the strong recommendations of statisticians to avoid dichotomizing continuous measures (e.g., Maxwell & Delaney, 1993). However, in many developmental domains it is quite difficult to avoid using categorical measures even when the researcher suspects that the underlying developmental process may be continuous. For example, research on children's theory of mind (see Bartsch & Wellman, 1995, for a review) often uses search tasks in which children judge an object to be in a single location, a clearly categorical measure, even though the children's knowledge may not be categorical. Work on the A-not-B error (e.g., Marcovitch & Zelazo, 1999) uses similar search methodology. Research that involves interviewing young children about their knowledge and beliefs (e.g., Flavell, Green, Flavell, & Lin, 1999; Woolley, Phelps, Davis, & Mandell, 1999) necessarily involves categorizing responses; constructing continuous measures would be extremely difficult. Similarly, research on problem solving focuses, in part, on the use of different types of strategies, an inherently categorical measure (e.g., Chen & Klahr, 1999; Dixon & Moore, 1996). Different categories of infant attachment have important implications for developmental outcomes, although these categories may not capture the full range of attachment (e.g., Vondra & Barnett, 1999). Similarly, the categorization of school-age children as peer rejected (e.g., Coie & Cillessen, 1993) has important implications for social development. Categorical measures are an unavoidable reality in developmental research. Further, although continuous measures have many highly desirable properties, they have surprisingly little advantage over categorical measures in terms of testing developmental ordering unless some stringent conditions (i.e., interval scales) are met. We return to this point in the final section.

Assuming that a researcher uses categorical measures, can he or she make useful statements about developmental ordering by examining the patterns in simple contingency tables? The answer to this question is yes; researchers can make useful and logically sound statements about developmental ordering hypotheses by examining the pattern of categorical measures across development even if the underlying variables develop continuously. But there are two important qualifications. First, the conclusions that can be drawn are quite different from the direct interpretations allowable when the underlying variables change discretely, or in a saltatory manner, from absent to present. Second, researchers need to apply the disconfirmatory logic of strong inference (Platt, 1964) to eliminate particular developmental ordering hypotheses. That is, because there is not a one-to-one relationship between observable patterns and underlying developmental orderings, researchers must shift their focus from confirming a single hypothesis to disconfirming competing hypotheses.

Recall that using a categorical measure of a continuous underlying variable imposes a cutoff on the variable. As shown in Figure 3, Skill A undergoes continuous development, but only people above a certain level are classified by the categorical measure as having the skill. The bar on the right represents the development of the underlying variable, Skill A. Darker areas of the bar represent higher developmental levels. All changes in Skill A occur within the period marked by the rectangle. Each line

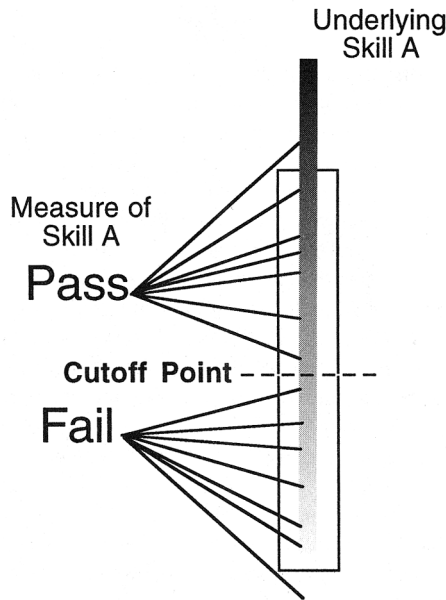


Figure 3. The relationship between the underlying variable, Skill A, and the dichotomous measure of Skill A is depicted. The bar on the right shows the continuous development of Skill A. Darker areas of the bar represent higher development levels. Each person will be classified as either passing or failing the measure of the skill, shown on the left. The lines connecting the bar to the measure show how the measure is related to the underlying variable. People with development that is greater (i.e., higher) than the cutoff point, shown by the dashed line, are classified as passing. That is, they are considered to have the skill. Those whose development is below the cutoff fail. They are considered not to have the skill.

connecting the bar to the measure on the left shows how the measure of Skill A would classify a person functioning at that level (i.e., pass or fail). People who are above the cutoff are classified as having Skill A (pass); people who score below the cutoff are classified as not having Skill A (fail). The cutoff, represented by the dashed line, is an aspect of the measure, and its placement relative to the development of the underlying variable is unknown in general.

The logical problem with interpreting developmental order from categorical measures of continuously developing variables is illustrated in Figure 4. The top panel of Figure 4 shows the contingency table relationship normally interpreted as developmental priority of Skill A over Skill B. Assuming that the underlying variables develop continuously, which developmental ordering hypotheses can explain the observed pattern? Surprisingly, developmental synchrony, both types of partial priority, as well as complete developmental priority of Skill A over Skill B can account for the observed pattern in the contingency table. We show this with the lower panels of Figure 4. Each panel has two cutoff points—one for each variable—shown as dashed lines. People scoring higher than the cutoff are classified as having the skill (i.e., Y); those scoring lower than the cutoff are classified as not having the skill (i.e., N). Therefore, the quadrants formed by the two intersecting cutoff points directly correspond to the quadrants in the observed contingency table. If the curve runs through the quadrant, people will be found in the corresponding cell in the contingency table.

It can be seen in Figure 4 that each of the four panels in the middle two rows can produce the pattern shown in the contingency

table. As expected, the complete and partial priority hypotheses that specify Skill A developing before Skill B can produce the observed pattern. However, other underlying patterns of continuous development can also produce the observed pattern. For instance, the synchrony relationship is also consistent with the observed pattern, and, perhaps more disturbingly, the reverse partial priority hypothesis (Skill B develops more quickly than Skill A)

		Skill B	
		N	Y
Skill A	Y	X	X
	N	X	O

Potential Underlying Relationships:

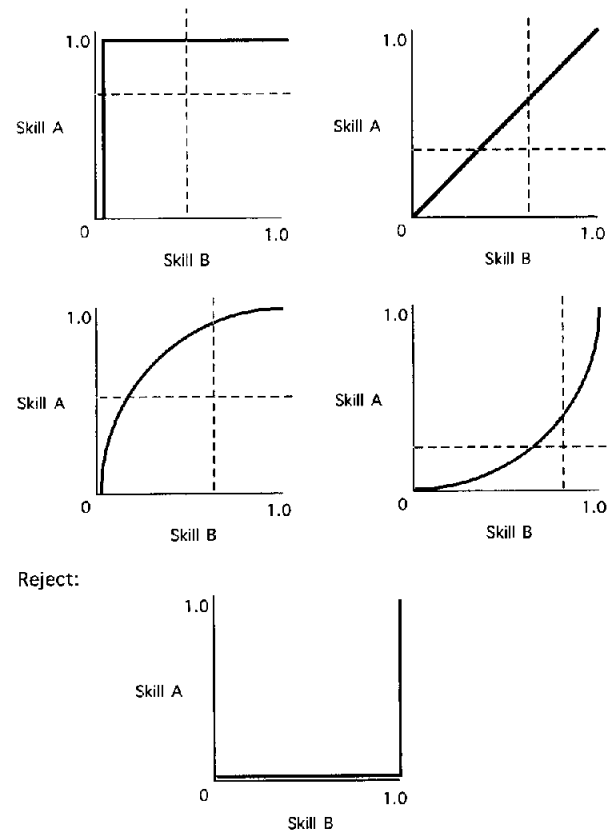


Figure 4. The contingency table in the top row shows the pattern usually interpreted as developmental priority of Skill A over Skill B (Y = yes [having the skill]; N = no [not having the skill]; X = people present in cell; O = people not present in cell). The middle two panels show the potential underlying relationships (i.e., developmental ordering hypotheses) that are consistent with the pattern in the contingency table. The dashed lines indicate the placement of the cutoff point imposed by each dichotomous measure. The quadrants formed by the dashed lines correspond directly to the cells in the contingency table. If the curve passes through the quadrant, people will be in the cell. The bottom row shows the opposite complete priority hypothesis—Skill B begins and completes development before Skill A—which can be rejected when the contingency table pattern shown in the top row is observed.

can also produce the observed pattern. Importantly, the reverse complete priority hypothesis (Skill B completes development before Skill A begins to develop), shown in the bottom panel, cannot explain the observed pattern. Regardless of how the cutoff points are drawn, this underlying relationship cannot produce the pattern shown in the contingency table. Therefore, when the pattern of data normally interpreted as developmental priority of Skill A over Skill B is observed, researchers can confidently *reject* the opposite complete priority hypothesis. However, further conclusions are not possible if the underlying variables undergo continuous development.

The top panel of Figure 5 shows the pattern of data normally interpreted as developmental synchrony, or simultaneous development of two skills. The middle two rows show developmental ordering hypotheses that are consistent with this observed pattern. Developmental synchrony of the underlying variables, of course, can produce this pattern in a contingency table. Likewise, both partial priority hypotheses can also produce the observed pattern. However, neither of the complete priority hypotheses (bottom row) can explain the observed pattern. That is, regardless of how the cutoff points are placed, neither of these underlying relationships can produce the pattern in the contingency table. Therefore, both complete priority hypotheses may be rejected, but it is important to remember that both partial priority hypotheses must be retained.

The relationship between observable data patterns and developmental ordering hypotheses (i.e., synchrony, partial priority, and complete priority) is summarized in Table 1. The top two rows show the allowable conclusions for categorical measures when development is assumed to be discrete (top row) or continuous (second row), as demonstrated in the discussion above. Thus far, we have focused on simple  $2 \times 2$  contingency tables for three reasons. First, as mentioned above, dichotomous categorical measures are used in developmental psychology whenever a pass-fail criterion for scoring a task is used. Second, dichotomous measures provide the clearest example of the logical issues. Third, in terms of testing developmental ordering hypotheses, very little is gained by having multivalued, as opposed to dichotomous, categorical measures.

Categorical measures can, of course, have more than two values. For example, understanding of a false belief task could be measured as *no understanding*, *poor understanding*, *fair understanding*, or *complete understanding* depending on the child's ability to explain his or her reasoning. As categorical measures take on many values, the measures begin to function increasingly like ordinal-level continuous measures. Although this may, at first, seem to impose a strong constraint on the relationship between the measure and the underlying variable, having an ordinal measure does not allow for stronger conclusions about developmental ordering. In fact, unless the continuous measures can be shown to have the equal-interval property, they do not offer any logical advantage over categorical measures for testing developmental order. The issue with continuous measures is that the shape of the observed curve, which can be seen by examining the relationship between the measures, does not necessarily reflect the shape of the underlying relationship unless the measures have the equal-interval property (Dixon, 1998). Psychological research rarely meets the conditions for establishing an equal-interval scale (Birnbau, 1982a, 1982b; Krantz, Luce, Suppes, & Tversky, 1971; Michell, 1990).<sup>5</sup>

The bottom two rows of Table 1 show the relationship between observable data patterns and developmental ordering hypotheses when a researcher uses continuous measures. Two data patterns, curvilinear and linear, are relevant to developmental ordering

when two continuous measures are cast in a scatterplot. These patterns are directly analogous to patterns in a contingency table, as shown in our discussion of Figure 2. A curvilinear pattern has traditionally been taken as evidence of developmental priority (e.g., Skill A develops before Skill B). A linear pattern has traditionally been taken as evidence of developmental synchrony (e.g., Skills A and B develop at the same time).

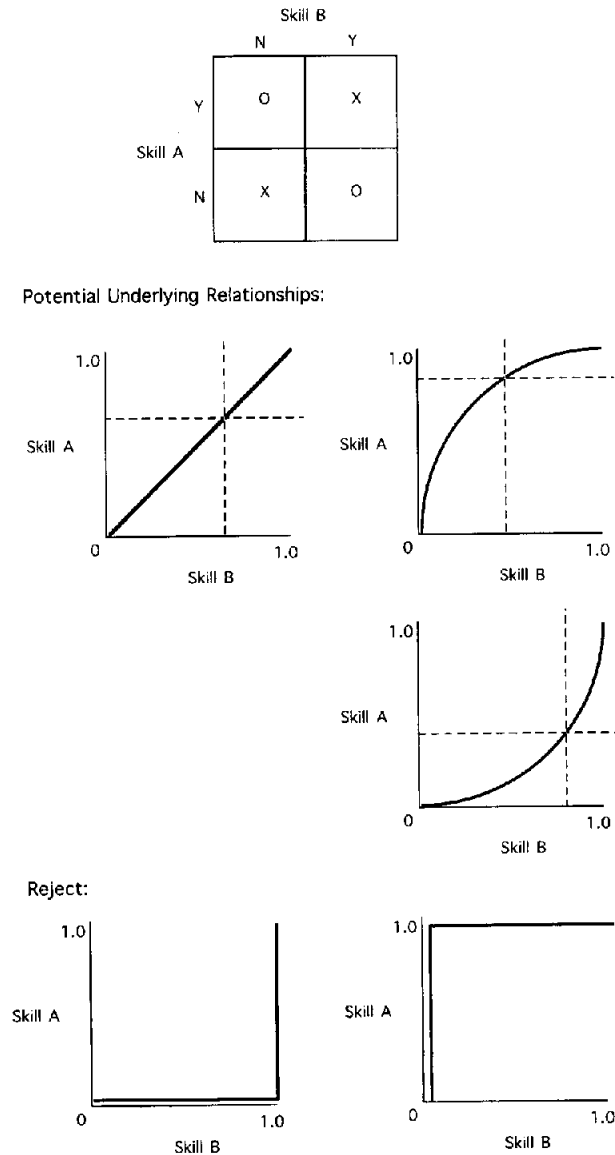
Note that when the underlying variables undergo continuous development, the relationship between the observed patterns and the viability of the hypotheses is exactly the same for categorical measures and ordinal-scale continuous measures. Observing a particular pattern with ordinal or categorical measures disconfirms the same hypotheses if the underlying variables develop continuously. If two interval-scale continuous measures are used, more hypotheses can be disconfirmed, because the shape of a scatterplot or curve reflects the shape of the relationship between the underlying variables. However, establishing interval scales is quite difficult. (See Dixon, 1998, for a more complete discussion of ordinal and interval scales with respect to developmental ordering, and see Surber, 1984, for a discussion of some other aspects of ordinal and interval scales in developmental research.) One surprising implication of this analysis is that without interval-scale data, categorical and continuous (ordinal-level) measures provide equally effective evidence regarding developmental ordering.

## Conclusions and Implications

The interpretation of data collected to test developmental order hypotheses is straightforward when researchers use categorical measures of underlying variables that undergo discrete, or saltatory, development. The observed pattern then reflects the underlying relationship between the constructs. When the underlying variables develop continuously but the measures are categorical, researchers can still draw strong conclusions from data patterns cast in contingency tables. However, these conclusions are quite different from the normal interpretations of these patterns of data. Each pattern of data effectively disconfirms one or more of the competing developmental ordering hypotheses. However, it is equally important for researchers to realize that each pattern is also consistent with multiple ordering hypotheses.

The issue of discrete versus continuous development has long been recognized as a substantive theoretical issue in developmental psychology, but its impact on the interpretation of evidence regarding developmental order has rarely been addressed (see Wohlwill, 1973, pp. 58–79, for a notable exception). Flavell (1971) discussed the problem briefly in his classic paper on the stage concept. The current discussion importantly extends Flavell's original argument, showing that although a confirmatory strategy is not appropriate with categorical measures of continuous

<sup>5</sup> Birnbaum (1982b) explained, "It used to be said that 'measurement consists of assigning numbers to objects (or attributes thereof) according to rules.' . . . This view has now given way to the idea that measurement consists of the construction of homomorphisms between relational structures, where one structure represents an empirical domain with experimental operations and the other structure is mathematical" (p. 40). Michell (1990, p. 65), also following the approach of Krantz et al. (1971), outlined nine conditions that must be tested in a relational data structure in order to have measurements on an equal-interval or ratio scale. These conditions are not normally assessed in developmental research.



*Figure 5.* The contingency table in the top row shows the pattern usually interpreted as developmental synchrony of Skill A and Skill B (Y = yes [having the skill]; N = no [not having the skill]; X = people present in cell; O = people not present in cell). The middle two panels show the potential underlying relationships (i.e., developmental ordering hypotheses) that are consistent with the pattern in the contingency table. The dashed lines indicate the placement of the cutoff point imposed by each dichotomous measure. The quadrants formed by the dashed lines correspond directly to the cells in the contingency table. If the curve passes through the quadrant, people will be in the cell. The bottom row shows the two complete priority hypotheses. These hypotheses can be rejected when the contingency table pattern shown in the top row is observed.

change, researchers can effectively adopt a disconfirmatory strategy. To use a disconfirmatory strategy, researchers must generate a set of competing hypotheses and recognize how the data patterns they might observe bear on those hypotheses. One advantage of this approach is that the researcher must explicitly define what is meant by the hypothesis that one skill develops before another. An additional advantage of the disconfirmatory strategy is that it

prevents the overinterpretation of a data pattern that is consistent with a favored hypothesis; by generating a set of competing hypotheses at the outset, the disconfirmatory strategy explicitly shows whether or not more than one hypothesis remains viable after the data have been collected.

For the sake of simplicity, the presentation here has assumed simple repeated measures but cross-sectional designs. The basic logic of testing developmental order that we have outlined is not altered by using longitudinal or more complex hybrid designs (e.g., longitudinal-sequential, Achenbach, 1978). The logical implications of measures that impose cutoff points on continuous underlying variables are the same regardless of whether different individuals of different ages are measured or the same individuals are measured at different points in time. This can easily be seen by examining Figure 4. Suppose that a group of children is followed longitudinally across three testing sessions. At Time 1, they show no evidence of either Skill A or Skill B and, therefore, are in the lower left cell of the  $2 \times 2$  table at the top of Figure 4. At Time 2, these same children show evidence of Skill A but not Skill B (i.e., the upper left cell of the table). At Time 3, they show evidence of both skills (i.e., the upper right cell of the table).

Next examine the five developmental ordering hypotheses graphically represented in Figure 4 below the  $2 \times 2$  table. In the cross-sectional case, individuals are sampled from different points along the developmental trajectory. In the longitudinal case, individuals are tested repeatedly as they move along the developmental trajectory. All four hypotheses that are viable in the cross-sectional situation are also viable with a longitudinal design. For example, the reverse partial priority hypothesis—Skill B develops more quickly than Skill A (shown in the third row on the right side of Figure 4)—is just as viable with a longitudinal design as with a cross-sectional design. Children are first classified as not having either Skill A or Skill B, because they fall below both cutoff points. Next, children are classified as having Skill A, because they exceed its low cutoff point, but not Skill B. Finally, they are classified as having both skills. Likewise, the opposite complete priority hypothesis that is rejected with a cross-sectional design is rejected with a longitudinal design as well. In summary, longitudinal designs do not confer any advantage (or disadvantage) over cross-sectional designs in terms of the measurement issues that are the focus here.

Our presentation of the logical issues in testing developmental sequences has assumed that the observed data patterns are statistically reliable. Some excellent methods exist for testing the reliability of the pattern of data in a contingency table. A very accessible introduction to these methods that focuses directly on developmental ordering can be found in Froman and Hubert (1980). Other sources include Hildebrand et al. (1977) and Wickens (1998). As Froman and Hubert (1980) explained, one method, prediction analysis, allows one to specify cells in the table that should be empty, according to the developmental ordering model, and to test the hypothesis of statistical independence for that model. Models may also be compared relative to one another. We emphasize that the appropriate interpretation of results from such analyses should be based on the disconfirmatory logic outlined above. That is, the statistical analysis which establishes the reliability of the pattern in a table weighs directly against the developmental ordering hypotheses that conflict with that pattern.

Researchers testing developmental order will benefit from careful consideration of the nature of their measures and their assump-

Table 1  
*Relationship Between Observable Data Patterns and Developmental Ordering Hypotheses for Categorical and Continuous Measures of Developmental Change*

Type of measure	Type of developmental change	Observed data pattern	Disconfirmed hypotheses	Viable hypotheses
Categorical	Discrete	a > b	B > A, AB	A > B
		ab	A > B, B > A	AB
Categorical	Continuous	a > b	B > A	A > B, A ≥ B, AB, B ≥ A
		ab	A > B, B > A	AB, A ≥ B, B ≥ A
Ordinal	Continuous	a > b	B > A	A > B, A ≥ B, AB, B ≥ A
		ab	A > B, B > A	AB, A ≥ B, B ≥ A
Interval	Continuous	a > b	B > A, AB, B ≥ A	A > B, A ≥ B
		ab	All others	AB

Note. The notation a > b refers to the priority pattern usually taken as evidence that Skill A develops before Skill B. The notation ab refers to the synchrony pattern that is usually taken as evidence that Skills A and B develop together. The notation A > B refers to the complete priority hypothesis (Skill A over Skill B). The notation B > A refers to the opposite complete priority hypothesis. A ≥ B refers to the partial priority hypothesis (Skill A over Skill B); B ≥ A refers to the opposite partial priority hypothesis. AB refers to the synchrony hypothesis. The hypotheses are defined in the text and graphically represented in Figure 2.

tions about the underlying variables they are attempting to measure. All types of data do not bear equally on the question of developmental ordering. Depending on whether it is justifiable to treat the underlying variables as genuinely categorical (or as showing saltatory development), the allowable interpretations of the data patterns differ radically. A number of sophisticated methodologies are now available to address hypotheses about continuous versus saltatory development (e.g., Brainerd, 1979; Thomas & Lohaus, 1993; Thomas, Lohaus, & Kessler, 1999; van der Maas & Molenaar, 1992; see Brainerd, 1993, for a very accessible review). Similarly, interval-scale continuous measures allow for strong conclusions about developmental ordering. Establishing that measures reasonably approximate an equal-interval scale is an arduous but quite possible task (Anderson, 1976; Dixon, 1998). Understanding how the observed pattern is constrained by, but is not a direct reflection of, the underlying relationship is critical if researchers are to draw logically sound conclusions from their data.

That said, it is important to acknowledge that in many areas of developmental research it will be exceedingly difficult to demonstrate either (a) that the underlying variables are categorical or (b) that the continuous measures approximate interval scales. For example, because infants tend to produce very noisy data, it will be difficult for researchers investigating infant cognition to establish whether their underlying variables are categorical. Researchers in these areas will be unable to eliminate key alternative hypotheses, such as the opposite partial priority hypothesis, on logical grounds. Instead, they will have to rely on converging evidence, strong relations between theory and data, and other extra-evidentiary factors to build convincing arguments about particular developmental ordering hypotheses. Therefore, one important implication of the current analysis is that when researchers are faced with measures of unknown (and perhaps unknowable) properties, strong conclusions about developmental ordering will require the careful integration of results across multiple studies that address a single hypothesis in different ways. Apparently conflicting results would be just as informative as mutually supporting results, because both types of results would help narrow the range of potential hypotheses. For example, suppose one researcher observes the priority pattern usually interpreted as Skill A develops before Skill B in a

2 × 2 contingency table. A second researcher observes the opposite priority pattern. On the surface, these results appear to be in direct conflict with each other. But, in fact, taken together they provide considerable information about the underlying relationship between Skills A and B; specifically, both complete priority hypotheses are disconfirmed.

Philosophers of science have disagreed about how prominent the role of disconfirmation should be in scientific inquiry (cf. Carnap, 1956; Feigl, 1976; Popper, 1959; Quine & Ullian, 1978). However, virtually all philosophies of science would agree that weakly grounded confirmatory conclusions should not be accepted and that an important way that science advances is by making alternative hypotheses explicit and testable. The logical analysis of developmental priority presented here should aid the field in both latter respects.

## References

- Achenbach, T. M. (1978). *Research in developmental psychology: Concepts, strategies, methods*. New York: Free Press.
- Anderson, N. H. (1976). How functional measurement can yield validated interval scales of mental quantities. *Journal of Applied Psychology, 61*, 677-692.
- Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York: Oxford University Press.
- Birnbaum, M. H. (1982a). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401-485). Hillsdale, NJ: Erlbaum.
- Birnbaum, M. H. (1982b). Problems with so-called "direct" scaling. In J. T. Kuznicki, R. A. Johnson, & A. F. Rutkiewicz (Eds.), *Selected sensory methods: Problems and approaches to hedonics, ASTM STP 773* (pp. 34-48). New York: American Society for Testing and Materials.
- Brainerd, C. J. (1977). Response criteria in concept development research. *Child Development, 48*, 360-366.
- Brainerd, C. J. (1979). Markovian interpretations of conservation learning. *Psychological Review, 86*, 181-213.
- Brainerd, C. J. (1993). Cognitive development is abrupt (but not stage-like). *Monographs of the Society for Research in Child Development, 58*(9, Serial No. 237), pp. 170-190.
- Brainerd, C. J., & Hooper, F. H. (1975). A methodological analysis of



- developmental studies of identity conservation and equivalence conservation. *Psychological Bulletin*, 82, 725-737.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Carnap, R. (1956). The methodological character of theoretical concepts. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science* (Vol 1). Minneapolis: University of Minnesota Press.
- Case, R., & Okamoto, Y. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development*, 61(1-2, Serial No. 246).
- Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of cognitive deficits. *Psychological Bulletin*, 79, 380-385.
- Chapman, L. J., & Chapman, J. P. (1978). The measurement of differential deficit. *Journal of Psychiatric Research*, 14, 303-311.
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70, 1098-1120.
- Coie, J. D., & Cillessen, A. H. N. (1993). Peer rejection: Origins and effects on children's development. *Current Directions in Psychological Science*, 2, 89-92.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Cornell, E. H., Heth, C. D., & Alberts, D. M. (1994). Place recognition and way finding by children and adults. *Memory & Cognition*, 22, 633-643.
- Dixon, J. A. (1998). Developmental ordering, scale types, and strong inference. *Developmental Psychology*, 34, 131-145.
- Dixon, J. A., & Moore, C. F. (1996). The developmental role of intuitive principles in choosing mathematical strategies. *Developmental Psychology*, 32, 241-253.
- Feigl, H. (1976). Defense of orthodoxy. In M. Marx & F. Goodson (Eds.), *Theories in contemporary psychology* (2nd ed., pp. 165-177). New York: Macmillan.
- Flavell, J. H. (1971). Stage-related properties of cognitive development. *Cognitive Psychology*, 2, 421-453.
- Flavell, J. H., Green, F. L., Flavell, E. R., & Lin, N. T. (1999). Development of children's knowledge about unconsciousness. *Child Development*, 70, 396-412.
- Froman, T., & Hubert, L. J. (1980). Application of prediction analysis to developmental priority. *Psychological Bulletin*, 87, 136-146.
- Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9, 139-150.
- Hildebrand, D., Laing, M., & Rosenthal, A. (1977). *Prediction analysis of cross classifications*. New York: Wiley.
- Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin*, 109, 490-501.
- Kail, R. (1997). Processing time, imagery, and spatial memory. *Journal of Experimental Child Psychology*, 64, 67-78.
- Karmiloff-Smith, A. (1991). Beyond modularity: Innate constraints and developmental change. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 171-197). Hillsdale, NJ: Erlbaum.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement*. New York: Academic Press.
- Krathwohl, D. R. (1985). *Social and behavioral science research: A new framework for conceptualizing, implementing, and evaluating research studies*. San Francisco: Jossey-Bass.
- Lewis, M., Sullivan, M. W., Stranger, C., & Weiss, M. (1989). Self-development and self-conscious emotions. *Child Development*, 60, 146-156.
- Marcovitch, S., & Zelazo, P. D. (1999). The A-not-B error: Results from a logistic meta-analysis. *Child Development*, 6, 1297-1313.
- Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, 113, 181-190.
- Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NJ: Erlbaum.
- Newcombe, N., Huttenlocher, J., Drumme, A. B., & Wiley, J. G. (1998). The development of spatial location coding: Place learning and dead reckoning in the second and third years. *Cognitive Development*, 13, 185-200.
- Perner, J. (1988). Higher-order beliefs and intentions in children's understanding of social interaction. In J. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 271-294). New York: Cambridge University Press.
- Platt, J. R. (1964). Strong inference. *Science*, 146, 347-353.
- Popper, K. R. (1959). *The logic of scientific discovery*. New York: Harper.
- Quine, W. V. O., & Ullian, J. S. (1978). *The web of belief* (2nd ed.). New York: Random House.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, 7, 1-75.
- Simon, T. J., & Klahr, D. (1995). A computational theory of children's learning about number conservation. In T. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 315-353). Hillsdale, NJ: Erlbaum.
- Surber, C. F. (1984). Issues in using quantitative rating scales in developmental research. *Psychological Bulletin*, 95, 226-246.
- Thomas, H., & Lohaus, A. (1993). Modeling growth and individual differences in spatial tasks. *Monographs of the Society for Research in Child Development*, 58(9, Serial No. 237), pp. v-169.
- Thomas, H., Lohaus, A., & Kessler, T. (1999). Stability and change in longitudinal water-level task performance. *Developmental Psychology*, 35, 1024-1037.
- van der Maas, H. L. J., & Molenaar, P. C. M. (1992). Stagemwise cognitive development: An application of catastrophe theory. *Psychological Review*, 99, 395-417.
- Vondra, J. I., & Barnett, D. (1999). Atypical attachment in infancy and early childhood among children at developmental risk. *Monographs of the Society for Research in Child Development*, 64(3, Serial No. 258).
- Wickens, T. D. (1998). Categorical data analysis. *Annual Review of Psychology*, 48, 537-558.
- Wohlwill, J. F. (1973). *The study of behavioral development*. New York: Academic Press.
- Woolley, J. D., Phelps, K. E., Davis, D. L., & Mandell, D. J. (1999). Where theories meet magic: The development of children's beliefs about wishing. *Child Development*, 70, 571-587.
- Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, 30, 111-153.

Received March 6, 2000

Revision received June 20, 2000

Accepted June 26, 2000 ■