Strong Tests of Developmental Ordering Hypotheses: Integrating Evidence From the Second Moment

James A. Dixon
University of Connecticut

Developmental ordering is a fundamental prediction in developmental science. However, tests of ordering hypotheses are not generally available for continuously developing variables. One promising test of developmental ordering, the shape of the relationship between 2 variables, requires that changes in each underlying variable are captured equally well across the developmental span (measures are linearly related to the variables). If either measure is more sensitive to earlier or later developmental changes in the underlying variable, the shape of the relationship changes radically. The article demonstrates that the viable alternative hypotheses for an observed developmental relationship require specific types of nonlinearity in measurement and, therefore, have testable predictions for the residuals (the second moment). Ordering as evidence in developmental science is discussed.

Developmental hypotheses predict the order in which abilities, mental structures, skills, behaviors, processes, and a myriad of other theoretically important items emerge. This was certainly true of classical structural approaches such as Piaget’s, but it is equally true of research today. Indeed, it is likely always to be true of developmental science because developmental order allows us to assess fundamental predictions about the nature of the developing system, for example, whether the development of one aspect of the system is contingent on the development of another. The developmental ordering of items is also used descriptively in new research domains. In this case, the developmental ordering is an important phenomenon that must be explained theoretically in the new research area. It is difficult to imagine a more important type of evidence for developmental theory. Developmental theories should rise and fall on the developmental orderings they predict.

Consider a few examples of developmental ordering hypotheses from a broad range of content areas. In cognitive development, Gentner and her colleagues (Gentner & Medina, 1998; Rattermann & Gentner, 1998) have proposed that children’s processing of similarity shifts from object based to relation based. Cohen, Chaput, and Cashon (2002) proposed that infants’ learning in a domain follows a hierarchical developmental progression in which low-level information must be learned before higher order regularities can be represented. In theory-of-mind research, Wellman and colleagues (Bartsch & Wellman, 1995; Wellman, Cross, & Watson, 2001; Wellman & Woolley, 1990) proposed that children’s theory of mind initially consists of states of perception, emotion, and desire, and only later do children develop theories about internal representations. Carlson and Moses (2001) hypothesized that inhibitory control may be a developmental prerequisite for theory of mind. In language development, Golinkoff, Hirsh-Pasek, and Hollich (1999) proposed that early word learning depends on three principles: reference, extendibility, and object scope. These principles are precursors to later principles that further refine language development. In research on the self, Lewis and Ramsay (1997) suggested that frontal lobe development may be an important developmental precursor to self-recognition. In social development, Cillessen and Mayeux (2004) proposed that physical aggression is replaced by relational aggression as a means to achieve social dominance.

I intend these examples to illustrate the fact that developmental ordering hypotheses are a substantial part of developmental theorizing regardless of area. Because ordering hypotheses make a strong empirical prediction, they should be a crucial way of connecting developmental theory and data. If one were to peruse the developmental literature in search of developmental ordering hypotheses, one would find them in nearly every domain. However, one would

I thank Toon Cillessen and Colleen Moore for helpful conversations regarding the issues discussed in the manuscript.

Correspondence concerning this article should be addressed to James A. Dixon, Department of Psychology, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT 06269-1020. Electronic mail may be sent to james.dixon@uconn.edu.

© 2005 by the Society for Research in Child Development, Inc. All rights reserved. 0009-3920/2005/7601-0001
also find that researchers rarely address predictions about developmental ordering directly. The reason for this is simple. Many researchers, perhaps even most, assume that developmental change is continuous, rather than saltatory, and tests of developmental order for variables that develop continuously require fairly large longitudinal designs and interval scales. The purpose of this article is to present a general method for testing developmental order, when development is assumed to be continuous (and measured with continuous variables). I show that strong conclusions about observed developmental relationships can be made by drawing on information contained in the second moment, specifically, in the pattern of residuals.

**Fundamental Problems for Testing Developmental Ordering Hypotheses**

Strong tests of developmental ordering for continuous measures have been remarkably difficult to construct. The problem is straightforward. Because we assess different constructs using different measures, measures that may vary widely in their scale properties, comparing those measures in a meaningful way is difficult (Chapman & Chapman, 1973, 1978). For example, comparing the means of two measures (e.g., a measure of short-term memory and a measure of language comprehension) rarely provides any information about their developmental order. Our measures tell us about the relative standing of individuals along a particular dimension (e.g., language comprehension) but do not provide information about their relative standing across dimensions (e.g., whether one’s level of language comprehension is more developmentally advanced than one’s level of short-term memory).

Given that directly comparing two measures does not reveal information about their developmental ordering, researchers have often looked at the shape of the relationship between the two measures over the developmental span of interest (Froman & Hubert, 1980; Guttman, 1944; Wohlwill, 1973). Different developmental ordering hypotheses predict different relational patterns. For example, Kail (1997) proposed that a wide variety of cognitive tasks would develop in synchrony because they were strongly influenced by a single underlying developmental dimension, cognitive speed. This developmental synchrony hypothesis predicts that the relationship between these tasks should be linear; developmental increases in cognitive speed will yield improved performance on tasks that depend on cognitive speed, a prediction that is now supported by much data (see Kail, 2000, for a review). Similarly, Bates and Goodman (1999) argued that grammar emerges from the developing lexicon. Their hypothesis predicts that lexical development should precede grammatical development, resulting in a curvilinear pattern when a measure of grammatical complexity is plotted as a function of number of words in the lexicon. Bates and Goodman reviewed a considerable body of evidence that demonstrates this predicted relationship.

Examining the shape of the relationship circumvents the problem of interpreting the absolute level of performance on a measure but quickly runs into problems of its own. If one of the measures is more sensitive to early developmental changes than it is to later changes (or vice versa), inferences based on the shape of the relational pattern become problematic. I refer to this problem as nonlinear mapping between the underlying variable and its measure and discuss it as a type of ordinal scale, but I wish to emphasize here that it is a fundamental problem in developmental research. How can we make inferences about development if our measures are not equally responsive across the developmental range? I return to this problem in more detail after presenting a more concrete way of representing developmental ordering hypotheses and the associated measurement problems.

First, I describe in detail three types of developmental ordering hypotheses, present a way of representing those hypotheses graphically, and discuss the data pattern predicted by each. Second, I review the difficulty with interpreting those data patterns directly, given ordinal, nonlinear mappings between the underlying variable and the observable measure. Third, I present a different way of representing the problem caused by ordinal scales and show how to eliminate hypotheses about systematic nonlinear mapping using tests of homoscedasticity.

**Developmental Ordering Hypotheses**

Although the content of developmental ordering hypotheses varies from area to area, structurally, three general types of hypotheses are often of interest. One hypothesis, which I refer to as complete priority, proposes that one item starts and completes development before a second item begins development. Following Flavell (1971), I use the term *item* as a generic label for any developing variable one might wish to study. This hypothesis is shown graphically in the top panels (i and ii) of Figure 1. In each panel, development is represented vertically; an individual can be thought of as traveling from the bottom of the
Figure 1. Developmental ordering hypotheses and predicted data patterns. Panels i through v show developmental ordering hypotheses. Individuals can be thought of as traveling up the vertical line labeled “Development” at the center of each panel. The developmental status of an item is represented by the saturation of the bar; as an item develops, the bar becomes darker. The period of developmental change for each item is marked by a rectangle. The upper two panels (i and ii) show examples of the complete priority hypotheses. In Panel i, item A starts and completes development before item B begins development. Panel ii shows the opposite relationship: Item B starts and completes development before item A begins development. The middle panels (iii and iv) show the partial priority hypotheses. Panel iii shows that item A begins development before item B but does not complete development until after item B begins. Panel iv shows the opposite relationship, partial priority of item B relative to item A. The bottom panel (v) shows the synchrony relationship: Items A and B start and complete development at the same time. The complete and partial priority developmental ordering hypotheses predict curvilinear data patterns when one is plotted as a function of the other. Idealized examples of these data patterns (Panels x and xi) are presented next to the respective hypotheses. The synchrony hypothesis predicts a linear relationship when one item is plotted as a function of the other. An idealized example of this pattern is shown next to the synchrony hypothesis (Panel xii).
The line labeled “Development” to the top. Consider the top left panel (panel i), which shows the complete priority of item A over item B. Item A is represented by the bar on the left. It begins development at the bottom of the panel; developmental changes in the item are represented by increasing saturation of the bar. A rectangle around the bar marks the period during which developmental changes occur. The development of item B is represented similarly on the other side of the panel. Note that all changes in A occur before any changes in B. The top right panel (ii) shows the opposite hypothesis, complete priority of B over A.

A second type of hypothesis, partial priority, proposes that development of one item begins prior to the development of a second item. However, development of the first item is not completed until after the second item begins development. The middle panels (iii and iv) of Figure 1 show examples of this type of hypothesis graphically. For example, on the left side (panel iii), item A begins development prior to item B, but their developmental periods overlap.

Finally, the synchrony hypothesis proposes that the development of two items occurs simultaneously; they start and complete development at the same time and develop at the same rate. The lower panel (v) of Figure 1 shows this hypothesis graphically. Items A and B start development and complete development together. These three types of hypotheses—complete priority, partial priority, and synchrony—cover quite a bit of theoretical ground, but other hypotheses are, of course, possible. In general, two items can differ in terms of their onset (i.e., when they begin), completion (when they finish development), and rate of development (see Flavell, 1971). The method presented here can be easily extended to a broad range of hypotheses.

I assume that sampling has been done across the developmental periods of both items, which may be very short for some rapidly developing items (e.g., preference for faces; Turati, 2004) or protracted for items that emerge over many years (e.g., the ability to search visually the environment; Hommel, Li, & Li, 2004). To simplify the initial presentation of the approach, I first assume sampling has been done cross-sectionally. Longitudinal designs add some additional complexity because of the possibility of within-subject covariance, but the substantive measurement issues and solutions are the same. Later, I present a longitudinal example and discuss connections to some of the sophisticated longitudinal models that have been developed.

**Predicted Relationships and Observable Data Patterns**

The hypotheses presented in Figure 1 make predictions about the form of the relationship when one item is plotted relative to the other. Assume that individuals are sampled from along the developmental line. Each individual is at a particular point in development on each continuously developing item, and assessment of the items is done at the same time or as close in time as possible. Plotting each individual’s level on one item as a function of his or her level on the other yields information about the developmental relationship between the items.

The complete priority hypotheses predict that development of one item will be complete before the development of the second item begins. The idealized relationship would look like a step function, when one item is plotted as a function of the other. That is, the idealized plot would have a right angle marking the point at which one item completed development and the other began development. Panels vi and vii in Figure 1 show the step-function relationship for each complete priority hypothesis.

The partial priority hypotheses predict that one item begins to develop before another, that the second item then starts to develop before the first item is complete, and that the second item continues to develop after the first item is complete. The idealized plot would have three segments: one in which only the first item showed development, one in which both items showed development, and one in which only the second item showed development. Panels viii and ix in Figure 1 show these idealized relationships for each partial priority hypothesis.

Given the presence of measurement error and the inability to sample every point along the curve, it may be difficult to distinguish between analogous complete priority and partial priority hypotheses, even under good measurement conditions (an issue that is emphasized later). For now, note that at the observed level, we expect that both types of hypotheses will produce curvilinear data patterns, although the complete priority hypotheses predict a sharper bend. I represent this prediction by showing the idealized patterns connected with arrows to their respective curvilinear relationships (Panels x and xi in Figure 1). These plots are intended iconically; that is, they show the general form of the relationship (i.e., curvilinear and opening up or down) rather than the exact shape of the curve.

The synchrony hypothesis predicts that the two developing items will be linearly related. Developmental increments in one item should be accompa-
nied by developmental increments in the other item at a constant rate. This predicted data pattern is shown next to the synchrony relation in Figure 1 (Panel xii).

**Ordinal, Systematically Nonlinear Measures**

The different developmental ordering hypotheses make predictions about the shape of the relationship between the two developing items. However, testing these predictions relies crucially on the mapping between the underlying item and the measure of that item. Unfortunately, our measures are not direct reflections of the underlying variables. Rather, we assume that measures are systematically related to the underlying variables, but the nature of this relationship is rarely known (Cliff, 1993; Stevens, 1951; Surber, 1984). Consider, for example, the synchrony relationship presented again in the top panel of Figure 2 with the addition of two measures, one for each underlying variable. The measures are presented next to each developing item. Because the measures are the only observable result in the figure, it may help to consider them as on a separate plane coming off the page. The relationship between each measure and its respective underlying variable is shown by lines connecting the two. Following the line from the underlying variable to the measure gives the score for that level of development. (The actual numbers on the scale are, of course, arbitrary. We are concerned with patterns rather than absolute numerical values.) The measure of item A is linearly related to the underlying level of item A. Increases in the underlying variable A have the same effect on the measure, regardless of the level of A. That is, the measure indexes changes equally across the developmental range of A.

The situation for item B is different. The measure of item B is nonlinearly related to the underlying variable B. Early changes in underlying item B get very small changes in the measure, but later changes get large increases in the measure. The measure of item B is an ordinal scale, but it is systematically nonlinear. Although this might initially seem like an unlikely scenario, perhaps even requiring a conspiracy by nature against one’s research, consider that all such a nonlinear mapping indicates is that the measure is not equally responsive across the developmental range. For example, a measure of short-term memory may be particularly responsive to early developmental changes but relatively less responsive to later changes (or vice versa). Researchers usually hope that their measures are capable of picking up developmental changes across the broad range of development, but rarely would we assert that his or her measure of an item captures all changes with equal sensitivity. Therefore, ordinal, nonlinear scales are probably the modal scenario for developmental research rather than a bizarre or remote possibility.
The problem that arises, given a nonlinear mapping between a measure and the underlying variable, is that the shape of the observed relationship changes radically. For example, the underlying relationship between items A and B in Figure 2 is synchrony and should, therefore, produce a linear relationship. However, because item B is nonlinarily related to its measure, the observed relationship between A and B is curvilinear. The bottom panel of Figure 2 shows the shape of the observed relationship. Taken at face value, this observed relationship implies, erroneously, that item A develops before item B.

Dixon (1998) presented an analysis of the nonlinear mapping problem that showed that the data patterns usually taken to imply developmental ordering eliminate some ordering hypotheses. However, as the preceding discussion suggests, for each developmental ordering data pattern, multiple hypotheses about the underlying developmental relationship remain viable. Figures 3 and 4 show this situation graphically. In Figure 3, the observed priority data pattern, A prior to B, is shown at the top. Consistent with the usual interpretation of this pattern, both the complete and partial, A-prior-to-B, priority relationships can produce this pattern. These relationships are labeled as the “Standard” interpretations in the figure. However, given nonlinear mapping between measures and variables, the partial priority, B-prior-to-A, and the synchrony relationships can also produce this observed pattern. These relationships are labeled as the “Alternative” interpretations in the figure. Examples of nonlinear mappings capable of producing these results are shown in the figure. The nonlinear mapping shown for the synchrony hypothesis is analogous to that discussed for Figure 2. The B-prior-to-A alternative hypothesis has a nonlinear mapping between the underlying variable and the measure for both items A and B. The measure of item A is more responsive to earlier versus later changes. The measure of item B is more responsive to later versus earlier changes.

The complete priority relationship, B over A, can be rejected. Regardless of how one (ordinally) maps the measure and the underlying item, the complete priority hypothesis cannot generate the observed pattern. All changes in B must occur before changes in A; therefore, observing a pattern in which changes in A precede changes in B allows us to reject this hypothesis.

Figure 4 shows analogous implications for the data pattern usually taken to imply synchrony. Both partial priority hypotheses can produce the synchrony data pattern if the mapping between the item and the measures is nonlinear. Hence, both partial priority patterns are labeled as “Alternative” in the figure. The complete priority hypotheses can be rejected; because all changes in one variable must occur at a single level of the other, no (ordinal) mappings from the variables to the measures could produce the observed linear relationship.

This leads us to the dismal conclusion that no observed data pattern allows one to reject either the partial priority hypotheses or the synchrony hypothesis. Only complete priority hypotheses can be disconfirmed empirically. Dixon (1998) suggested a possible solution to this dilemma in which researchers would work to establish interval-level scales as part of their research program (Anderson, 1976, 1981, 1982). Interval-level scales have a linear mapping between the underlying construct and the measure, which allows for more direct interpretation of the shape of relationships. Such scales, therefore, offer strong advantages for testing developmental ordering hypotheses. Although I believe that approach has considerable merit, establishing even approximate interval-level scales is a daunting task. Indeed, it may be nearly impossible to do in some research areas (e.g., infant development) because fairly large within-subject designs are required.

**Representing Developmental Ordering Hypotheses as Simple Functions**

I propose another approach that capitalizes on the nonlinearity in the problem and the unavoidable presence of error and individual variation. Consider a different way of representing the underlying synchrony relationship between A and B using a standard linear model: \( B_u = \alpha + \beta \cdot A_u \). This is the model invoked for ordinary least squares (OLS) regression, analysis of variance, and other standard techniques; \( \alpha \) is the intercept, \( \beta \) is the slope, and \( A_u \) and \( B_u \) represent underlying items A and B. The equation posits that \( A_u \) is linearly related to \( B_u \) and is, therefore, another way of representing the synchrony relationship.

The model also includes an additional term, \( \varepsilon \), to represent error. Error, of course, is attributed to unintended differences in procedure and measurement, as well as to individual differences and the effects of unmeasured influences. Therefore, error enters into the model at both the measured and underlying levels. With standard OLS techniques, we usually do not make fine distinctions about the sources of error at the analytic level because regardless of the source, error is estimated with a single parameter, \( \varepsilon \). At the design level, however, researchers are well aware of
Figure 3. The implications of the A-prior-to-B data pattern. The top panel shows the data pattern usually interpreted as developmental priority of item A relative to item B. The panels in the remaining three rows show the various developmental ordering hypotheses. The A-prior-to-B, complete and partial priority hypotheses, shown on the left side (rows 2 and 3), are consistent with the standard interpretation of this data pattern and are therefore labeled as “Standard” in the figure. The B-prior-to-A, complete priority hypothesis cannot explain the observed relationship and can therefore be rejected (row 2, right side). Both the B-prior-to-A, partial priority and the synchrony hypotheses can explain the observed relationship, given a nonlinear mapping between at least one underlying variable and its measure, and are therefore labeled as “Alternative” hypotheses.
Data Pattern Interpreted as Developmental Synchrony of A and B

Figure 4. The implications of the synchrony data pattern. The top panel shows the data pattern usually interpreted as developmental synchrony of items A and B. The panels in the remaining three rows show the various developmental ordering hypotheses. The synchrony hypothesis, shown in the bottom panel, is the standard interpretation of the observed pattern. The complete priority hypotheses, shown in the second row, cannot explain the observed relationship and therefore can be rejected. Both partial priority hypotheses (third row) can explain the observed relationship, given a nonlinear mapping between at least one underlying variable and its measure, and are therefore labeled as “Alternative” hypotheses.
the potential for error to enter in a variety ways (Keppel & Wickens, 2004; see also Busemeyer, 1980). For example, decreasing the heterogeneity of the sample, tightening experimental procedures, and adding relevant covariates all decrease the error term at different levels. Advanced statistical techniques such as multilevel modeling explicitly represent error at multiple levels (Raudenbush & Bryk, 2002; Singer & Willett, 2003). Therefore, the model for synchrony is represented as: $B_u = \alpha + \beta * A_u + e_{uu}$ where $e_{uu}$ represents the effect of error at the level of the underlying variables. The error term is usually assumed to be drawn at random from a normal distribution with a mean of 0 and unknown variance. It is also assumed to be unrelated to the level of $A_u$. This latter assumption, called homoscedasticity, plays a key role in the following analysis.

**Homoscedasticity, Heteroscedasticity, and Developmental Order**

Assume that, as specified by the equation, A and B develop synchronously and are, therefore, linearly related at the underlying level. Further assume that the measure of A is linear, but the measure of B is nonlinear in the manner shown in Figure 2; early changes in B get smaller changes in the measure relative to later changes. Note that the nonlinear mapping between the underlying item B ($B_u$) and the measure of B ($B_m$) can be represented mathematically as a simple nonlinear transformation, such as being raised to a positive power greater than 1: $B_m = B_u^2 + e_m$. The additional error term, $e_m$, reflects the presence of error at the level of taking the actual measurement of B. Recall that $B_u = \alpha + \beta * A_u$. Therefore, by substitution, we find that $B_m = (\alpha + \beta * A_u + e_u)^2 + e_m$. (Squaring the values of $B_u$ obviously increases them dramatically as well as changing the shape of the relationship with $A_u$. The change in magnitude does not affect the analysis presented here.) This nonlinear transformation is, of course, causing all of the trouble. However, the transformation also leaves a signature that can be readily detected. The nonlinear transformation creates a strong relationship between $A_u$ and the portion of the error term that has been transformed. That is, it creates systematic heteroscedasticity, a violation of homoscedasticity, such that the absolute value of the residual term is positively related to the value of $A_u$.

The reason for this, perhaps surprising, new relationship between the error term and $A_u$ is easy to demonstrate informally. As the value of $A_u$ increases, the effect of the error term increases because the sum of both terms is squared. For example, consider the effect of increasing the error term, $e_{uu}$, by 2 units, for different values of the structural part of the equation, $\alpha + \beta * A_u$. If $\alpha + \beta * A_u = 8$ and $e_u = 2$, then $B_m = 100$ (ignoring the additive term $e_m$ for the moment). If $e_u$ increases to 4 and $\alpha + \beta * A_u$ remains constant (i.e., $\alpha + \beta * A_u = 8$), $B_m$ goes from 196 to 256, a change of 60 units. Therefore, the estimated effect of error will become strongly correlated with the level of $A_u$ and to the value of $A_{uu}$, its measured value.

This basic fact, that nonlinear transformations create systematic heteroscedasticity, allows us to evaluate whether any observed developmental ordering pattern is the result of nonlinear mapping between underlying variables and measures as opposed to being a reflection of the underlying relationship between the variables. To make this point more concretely, I next present a brief demonstration of how nonlinear mapping creates heteroscedasticity from an originally homoscedastic situation using data generated under standard (fixed-effect) linear model assumptions and an underlying synchrony relationship. A secondary aspect of the signature left by nonlinear mappings will also be revealed: a curvilinear pattern in the residuals.

**Demonstration of Heteroscedasticity From Homoscedasticity and Nonlinearity**

The general strategy here is simple. I first created two underlying variables that are linearly related and therefore consistent with the synchrony relationship. Next, I created variables to simulate measures of the underlying variables; one of these measures is a linear function of its underlying variable and the other is a nonlinear function of its underlying variable. This allows me to demonstrate how a known underlying relationship (i.e., synchrony) can produce a different observed relationship (i.e., priority). It also shows that the nonlinear mapping is detectable at the level of residuals; systematic heteroscedasticity results from the nonlinear transformation.

First, I created an underlying variable, $A_u$, with values that ranged from 1 to 100. $B_u$ was then created as linear function of $A_u$, and a normally distributed error term, $e_{uu}$, with a mean of 0 and a variance of 20: $B_u = \alpha + \beta * A_u + e_{uu}$. The intercept, $\alpha$, and the slope, $\beta$, were arbitrarily set at 5 and 1.5, respectively. The data set contained 1,000 cases generated by this simple model. The upper left panel of Figure 5 (Panel i) shows the scatter plot of values for $A_u$ and $B_u$. As
expected, if $A_u$ is used as a predictor of $B_u$, the fit is extremely good. The middle left panel (ii) of Figure 5 shows the pattern of residuals (observed – predicted), when $A_u$ is used to predict $B_u$, plotted as function of predicted values of $B_u$. The lower left panel (iii) shows the pattern of residuals, when $B_u$ is used to predict $A_u$, plotted as a function of predicted values of $A_u$. The right-side panels show the results for the measures of $A$ and $B$. $A_m$ is a linear function of $A_u$; $B_m$ is a nonlinear function of $B_u$. The middle right panel (v) shows the pattern of residuals (observed – predicted), when $A_m$ is used to predict $B_m$, plotted as a function of predicted values of $B_m$. The lower right panel (vi) shows the pattern of residuals, when $B_m$ is used to predict $A_m$, plotted as a function of predicted values of $A_m$.

Figure 5. Consequences of nonlinear mapping for the observed data pattern and residuals: Cross-Sectional model. The left-side panels show the results for the underlying variables, $A$ and $B$. The top left panel (i) shows the relationship between item $A$ and item $B$ at the underlying level. The middle left panel (ii) shows the pattern of residuals (observed – predicted), when $A_u$ is used to predict $B_u$, plotted as function of predicted values of $B_u$. The lower left panel (iii) shows the pattern of residuals, when $B_u$ is used to predict $A_u$, plotted as a function of predicted values of $A_u$. The right-side panels show the results for the measures of $A$ and $B$. $A_m$ is a linear function of $A_u$; $B_m$ is a nonlinear function of $B_u$. The middle right panel (v) shows the pattern of residuals (observed – predicted), when $A_m$ is used to predict $B_m$, plotted as a function of predicted values of $B_m$. The lower right panel (vi) shows the pattern of residuals, when $B_m$ is used to predict $A_m$, plotted as a function of predicted values of $A_m$. $A_u$ can be used as the dependent variable. The bottom left panel (iii) shows the residual values of $A_u$, as a function of predicted values of $A_u$. These latter two plots are a standard way of examining the pattern of the residuals (e.g., Cohen, Cohen, West, & Aiken,
2003; Pedhazur, 1982). Note that the pattern of residuals for both $B_m$ and $A_m$ are evenly distributed across their respective predicted values. The patterns show strong homoscedasticity; there is no relationship between the value of the residuals and the predicted value of the variable. Of course, in practice we never have access to this under-the-hood view of the relationship between the underlying variables; we only have measures of the variables in hand. I present these plots to give a concrete sense of the underlying relationship and the homoscedastic residuals. The plots also provide a point of comparison that allows us to see how a nonlinear mapping between a variable and its measure radically changes the shape of the underlying relationship and the homoscedastic residuals. The plots also provide a point of comparison that allows us to see how a nonlinear mapping between a variable and its measure radically changes the shape of the underlying relationship and the homoscedastic residuals.

Assume that the mapping between the underlying variable, $B_u$, and its measure, $B_m$, is nonlinear in the following way: The measure is more responsive to later versus earlier developmental changes. Consistent with the preceding discussion, I created $B_m$ by squaring $B_u$ and adding an additional error term, $e_{mb}$. $A_m$ was created by adding an error term to $A_u$: $A_m = A_u + e_{ma}$. Both of these normally distributed error terms, $e_{ma}$ and $e_{mb}$, have a mean of 0 and a variance of 20. The upper right panel of Figure 5 (Panel iv) shows a scatter plot of $A_m$ and $B_m$. The observed pattern is now curvilinear. (To facilitate comparison, I linearly rescaled $B_m$. The rescaling does not affect the pattern.) $B_m$ is strongly predicted by $A_m$ and $A_m^2$, the squared term captures the curvilinear nature of the relationship. (Adding the squared term is standard practice when the data pattern shows a single bend.) The middle right panel (v) of Figure 5 shows the residuals plotted again as a function of predicted values of $B_m$. Note that the pattern of residuals has changed radically compared with the plot next to it (in Panel ii). Although the residuals still have a mean of 0, the absolute value of the residuals is strongly and positively related to the predicted value of $B_m$.

When $B_m$ and $B_m^2$ are used to predict $A_m$, the pattern of residuals is different. There is a curvilinear relationship between the residual values and the predicted values of $A_m$, as can be seen in the lower right panel (vi) of Figure 5. Curvilinearity in the pattern of residuals usually indicates that one has incorrectly specified the form of the relationship between the variables, often by failing to include a nonlinear term. In this case, the curvilinearity in the residuals results from the fact that $B_m$ is a nonlinear function of $B_u$. Transforming it further to create the quadratic term, $B_m^2$, creates a predictor that is nonlinear relative to $B_u$ and, ultimately, to the dependent variable, $A_m$. As a result, when $B_m$ and $B_m^2$ are used as predictors, they systematically misspecify the form of the relationship. Although this curvilinear pattern can be dramatic, it only appears under some conditions (i.e., when the predictors and the underlying relationship have mismatching forms). Therefore, I consider it a secondary indicator of nonlinear mapping. In the next section, I show how particular developmental ordering hypotheses can be tested by examining the relationship between the two measures in conjunction with the pattern of residuals.

**Extending Tests of Homoscedasticity to Specific Alternative Ordering Hypotheses**

Recall that for each of the data patterns usually taken to imply developmental order, there are alternative hypotheses that were also capable of explaining the pattern of results. Each alternative hypothesis specifies a particular underlying developmental relationship, one that is qualitatively different from the standard interpretation of the data pattern. We would very much like to know whether one of these alternatives is driving the observed relationship or if we are entitled to the standard interpretation. Examining the residuals for evidence of nonlinear mappings helps in this regard. To develop this point, I again present the alternative hypotheses that can explain the priority, $A$-prior-to-$B$ data pattern, this time showing the three types of nonlinear mapping situations that can produce the observed data, given each hypothesis. Identifying the possible nonlinear mappings allows us test the alternative hypotheses by examining whether the residuals show the patterns predicted by those mappings.

**Alternative Hypotheses for the Observed Priority Relationship**

The top panel of Figure 6 (Panel i) shows the data pattern usually interpreted as priority of $A$ over $B$. The pattern is consistent with both $A$-prior-to-$B$ hypotheses, partial and complete priority; these were the standard interpretations shown in Figure 3. The data pattern cannot be explained by the $B$-prior-to-$A$, complete priority hypothesis; we were able to reject it earlier without considering the residuals. The $B$-prior-to-$A$, partial priority hypothesis, shown in the second row of panels, can create the observed pattern through nonlinear mapping between the measures and underlying variables. The synchrony hypothesis, shown in the third row of panels, can also create this data pattern given systematic non-
linear mapping. Therefore, the partial priority (B-prior-to-A) and synchrony hypotheses are the alternatives with which we are concerned.

The nature of the nonlinear mappings capable of producing the observed data pattern are largely the same for the two hypotheses and can be described as being of three types. First, the measure of A, which according to the observed data pattern appears to be developing more rapidly, may be more responsive to earlier rather than later changes. Examples of this situation are shown in the left-hand panels (ii and iii) of Figure 6. Second, the measure of B, which appears to be developing more slowly, may be more responsive to later rather than earlier changes. The center panels (iv and v) of Figure 6 show examples of this situation. Finally, both A and B have nonlinear mappings between the underlying variables and the measures.

Figure 6. Alternative hypotheses for the A-prior-to-B data pattern. The top panel (i) shows the A-prior-to-B priority pattern; the two lower rows show the alternative hypotheses that can explain this observed pattern. The second row of panels shows the three types of nonlinear mapping situations that can produce this data pattern, given the underlying relationship of partial priority, B prior to A. The third row of panels shows the three types of nonlinear mapping situations that can produce the observed data pattern, given the underlying synchrony relationship. The left-side panels (ii and iii) show a nonlinear mapping between the underlying variable A and its measure; early changes get larger gains than later changes. The middle panels (iv and v) show a nonlinear mapping between the underlying variable B and its measure; early changes in B get smaller gains relative to later changes. The right-side panels (vi and vii) show the situation in which both A and B have nonlinear mappings between the underlying variables and the measures.
given that the observed pattern is a reversal of the partial priority relationship, as opposed to a “bending” of the synchrony relationship. (More formally, one can consider the mappings between the underlying variables and measures as functions that either go up or down the ladder of powers; Mosteller & Tukey, 1977). Therefore, I use the synchrony relationship as an example, noting that the more extreme nonlinear mappings required by the partial priority hypothesis would create qualitatively similar but more extreme heteroscedastic patterns.

If we accept the usual assumptions of the standard (fixed-effect) linear model, including that the errors are unrelated to the level of the predictors, these alternative hypotheses can be evaluated by examining the pattern of the residuals. Because each alternative hypothesis can only generate the observed data pattern through specific nonlinear mappings and each mapping has predictable consequences for the pattern of residuals, failure to observe these predicted patterns strongly counters the alternative hypothesis.

The predicted residual patterns are shown in Figure 7. When the nonlinear mapping is a decelerating function (i.e., the measure is more responsive to earlier rather than later developmental changes), the absolute value of the residuals will be negatively related to the predicted values of that variable. For example, the left panel in the second row of Figure 7 (Panel ii) again shows an underlying synchrony relationship with a measure of A that is nonlinear; it is more responsive to earlier developmental changes. This produces the priority data pattern shown in Figure 6 and reproduced at the top of Figure 7 (Panel i). When \( B_m \) and \( B^2_m \) are used to predict \( A_m \), the residuals will be strongly heteroscedastic. The lower left panel (iv) of Figure 7 shows the residuals plotted as a function of the predicted values of \( A_m \). As can be seen in the figure, the absolute values of the residuals are negatively related to the predicted values of \( A_m \).

When the nonlinear mapping is an accelerating function (i.e., the measure is more responsive to later rather than earlier developmental changes), the absolute values of the residuals will be positively related to the predicted values. The middle panel (v) of the second row in Figure 7 again shows the synchrony relationship, but this time the measure of B is nonlinear; it is more responsive to later developmental changes. This also can produce the priority data pattern shown at the top of the figure. When \( A_m \) and \( A^2_m \) are used to predict \( B_m \), the pattern of the residuals will again be strongly heteroscedastic. The middle panel (vi) of the third row in Figure 7 shows the residuals of \( B_m \) as a function of the predicted values. The absolute values of the residuals are strongly and positively related to the predicted values. Panels iii and vii also show heteroscedastic patterns to some extent, although less dramatically than their counterparts. I present them here for completeness.

Finally, when both A and B are nonlinearly mapped (as in Panel viii of Figure 7), the measure of A is more responsive to early changes and the measure of B is more responsive to later changes, heteroscedasticity is seen in both patterns of residuals. When A is used to predict B, the absolute values of the residuals are positively related to predicted values of B (Panel ix). When B is used to predict A, the absolute values of the residuals are negatively related to the predicted values of A (Panel x). The curvilinear nature of the relationship is also dramatic here and, when seen in conjunction with systematic changes in the absolute value of the residuals, offers additional evidence of nonlinear mapping between the measure and underlying variable.

Note that regardless of which alternative hypothesis we consider, synchrony or the opposite partial priority, B can only undergo a single type of nonlinear mapping if it is to create the observed data pattern (i.e., A prior to B). The measure of B must be an accelerating function of the underlying variable. Similarly, only one type of nonlinear mapping of A can help create the observed pattern. The measure of A must be a decelerating function of the underlying variable. Therefore, the alternative hypotheses make clear predictions at the level of the residuals. The residuals for the variable that appears to be developing more slowly, B in the current example, are predicted to be heteroscedastic such that the absolute values of the residuals are positively related to the predicted values. The residuals for the variable that appears to be developing more quickly, A in the current example, are predicted to be heteroscedastic such that their absolute values are negatively related to the predicted values. Therefore, if neither of the predicted residual patterns is observed, the alternative hypotheses can be rejected; the observed data pattern cannot be explained by other ordering hypotheses. Conversely, if either residual pattern is observed, the observed relationship should not be interpreted as evidence of priority.

**Alternative Hypotheses for the Observed Synchrony Relationship**

The top panel of Figure 8 (Panel i) shows the synchrony data pattern. Recall that neither complete priority hypothesis could explain this pattern; these
hypotheses were rejected without reference to the residuals (see Figure 4). However, both partial priority hypotheses can explain this data pattern and, therefore, are the alternative hypotheses considered here. The second row of panels (ii, iv, and vi) shows the B-prior-to-A, partial priority hypothesis with the three types of nonlinear mapping situations that can give rise to the observed data pattern. Nonlinear mappings for A, B, and both A and B are in the left, center, and right panels, respectively. The third row of panels (iii, v, vii) shows the A-prior-to-B, partial priority hypothesis. Again, the nonlinear mappings for measures of A, B, and both A and B are shown in the left, center, and right panels, respectively. Depending on whether the alternative hypothesis is B prior to A or A prior to B, the nonlinear mappings for A are either decelerating or accelerating functions. Similarly, the nonlinear mappings for B
are either accelerating or decelerating, depending on the hypothesized underlying relationship. Therefore, the alternative hypotheses predict that the absolute value of residuals of either A or B may be either positively or negatively related to their respective predicted values. (Because these are essentially the same patterns presented in Figure 7, I do not present them again here.) If neither item shows the predicted residual pattern, the alternative hypotheses can be rejected. However, if either variable shows one of the predicted residual patterns, the observed relation should not be taken to imply developmental synchrony.

**Statistical Tests of Homoscedasticity**

Several methods have been recommended for analyzing the pattern of residuals. Cohen et al. (2003)
discussed the modified Levene test for assessing whether the pattern of residuals shows a significant positive or negative relationship with the predicted values, the types of patterns with which we are most concerned here. This test has the advantage of being simple to conduct. The sample is divided into two groups, usually splitting it at the median predicted value of the dependent variable. Within each group, the median value of the residuals is calculated. Finally, the absolute difference between each residual and the respective group median is computed. These scores are then subjected to a $t$ test. Rejecting the null hypothesis implies that the residuals are heteroscedastic. Failure to reject the null hypothesis implies homoscedasticity, given sufficient power. Because the modified Levene test is a simple $t$ test, its power can be easily evaluated post hoc using the tables provided by Cohen (1988). Similarly, given reasonable suppositions regarding effect size, the necessary sample size for a particular level of power can be obtained a priori. Other methods are also available to test for particular patterns in the residuals (see Fox, 1991; Goodall, 1983; Greene, 1997). See Tabachnick and Fidell (2001) for examples of how to employ many of these tests using the major statistical packages, as well as a discussion of the conceptual issues.

**Importance of the Structural Terms**

Because my emphasis thus far has been on the implications of the residuals for testing hypotheses about nonlinear mapping, I have not discussed the importance of the structural terms within the model. If one has hypotheses about developmental ordering, it is, of course, important to test whether item $A$ predicts item $B$ (or vice versa) and the reliability of any observed bending in the data pattern. For cross-sectional designs, OLS regression allows us to test the data patterns that are of concern here (i.e., the curvilinear priority and linear synchrony patterns).

Using the power polynomial representation of the curve makes the developmental ordering models nested; synchrony is nested within priority because priority has one additional term (e.g., $A^2$). This is convenient for comparing models and computing power. For developmental ordering hypotheses, the squared term is of particular concern because it carries information about the curvilinear nature of the observed pattern. If the contribution of the squared term is significant, it provides evidence of a developmental priority pattern. If the contribution of the squared term is not significant, it is important to evaluate the power of the test. Power, in this case, gives an indication of how confident we should be in rejecting the priority data pattern. Cohen (1988) provided methods and tables for estimating power and sample size for these models. In general, because we are working within the standard OLS regression framework, the sample sizes usually employed for a correlational study should suffice. It is worth noting that although the polynomial representation is very useful, it may sometimes be beneficial to explore other approaches for representing curvilinear relationships (see Cohen et al., 2003).

To emphasize the effects of nonlinear mapping between the underlying and observed levels as clearly as possible, I created measures that were very reliable; the proportion of observed variance due to true variance was high. However, the approach does not require measures that are unusually reliable. In fact, unexplained individual differences at the underlying level contribute to heteroscedasticity if a nonlinear mapping occurs. These deviations are nonlinearly transformed and they, therefore, become related to the magnitude of the predicted values, as described earlier. Unexplained variation at the measurement level, on the other hand, is undesirable, as it usually is. Poorly constructed instruments, fluctuation in participants’ interest, and so forth decrease power and make inference more difficult. As mentioned earlier, the power of tests of homoscedasticity and structural terms can be assessed using standard methods.

**Extending the Approach to Longitudinal Data**

To simplify the initial presentation of the method, I first considered cross-sectional data. This allows for several convenient simplifying assumptions regarding the distribution of error, and cross-sectional data are often collected in developmental work. However, longitudinal data offer some distinct advantages for addressing questions of developmental ordering (Wohlwill, 1973). Achenbach (1978) and Baltes and Nesselroade (1979) provided summaries of the relative advantages of longitudinal designs in developmental research, including the ability to assess the developmental trajectories of individuals as opposed to groups (see McArdle & Hamagami, 1991, for a discussion of integrating cross-sectional and longitudinal approaches). Although longitudinal data have many strengths, they do not alleviate the measurement issue with which we have been concerned: the potential nonlinear mapping between the underlying variable and its measure. This can be easily seen by reexamining the upper panel of Figure 2.
Suppose that, as an individual moves from the bottom of the developmental dimension (the center arrow) toward the top, his or her level of item B is repeatedly assessed via the measure. Just as in the cross-sectional situation, early changes in the underlying level of item B will get relatively small changes in the measure of B. Later changes in the underlying level of B will get larger changes in the measure. In the longitudinal case, just as in the cross-sectional case, the measure of B is all that we have in hand; therefore, the mapping between the underlying variable and the measure has the same distorting effect on what is observed. By the same token, the nonlinear mapping for data collected longitudinally will create a set of systematic relationships between the error term (or terms, in most longitudinal models) and predicted values, just as it did in the cross-sectional situation.

Next, I illustrate this point with an example using the underlying synchrony relationship and a simple growth curve model, which is analogous in many ways to the OLS regression model presented for the cross-sectional case. The model used in this example makes some standard assumptions about the error covariance structure (Singer & Willett, 2003), as such models routinely do, but the general approach can be extended to other error covariance structures.

In the growth curve model, the synchrony relationship between items A and B is represented as two hierarchically related submodels. The first model specifies that the level of an individual at a particular measurement occasion, $j$, can be represented as: $B_{uij} = \gamma_{10} + \gamma_{00} * A_{uij} + \epsilon_{ij}$. The $i$ subscript indexes individuals and $j$ indexes measurement occasions. As in the OLS regression model I used to represent the synchrony relationship for cross-sectional data, $B_{uij}$ is linear function of $A_{ui}$, but now multiple measurement occasions are represented (the $j$s).

The other major difference here is that the intercept ($\gamma_{00}$) and slope ($\beta_{1i}$) parameters are now allowed to vary across individuals. This variation is captured in the second set of models. The model for the intercept is: $\gamma_{00} = \gamma_{00} + \zeta_{00} * A_{uij}$ where $\gamma_{00}$ is the population average value for the intercept and $\zeta_{00}$ is the deviation of an individual’s intercept from the average intercept. The residual term, $\zeta_{00i}$, allows each individual’s intercept in the first model, $\gamma_{00}$, to vary around the population average intercept, $\gamma_{00}$. The analogous model for the slope is: $\beta_{1i} = \gamma_{10} + \zeta_{10} * A_{uij}$ where $\gamma_{10}$ is the population average value for the slope and $\zeta_{10}$ is the deviation of an individual’s slope from the average slope. The residual term, $\zeta_{10i}$, allows the slope parameter in the first model, $\beta_{1i}$, to vary around the population average slope, $\gamma_{10}$.

If we substitute the models just specified for the intercept and slope into the first equation, we get $B_{uij} = (\gamma_{10} + \zeta_{10i}) * A_{uij} + \epsilon_{ij}$. For current purposes, it is useful to reorganize the model into its structural and stochastic portions, yielding $B_{uij} = \gamma_{00} + \gamma_{10} * A_{uij} + (\epsilon_{ij} + \zeta_{00i} + \zeta_{10i} * A_{uij})$. Considered this way, the current model is closely analogous to the cross-sectional model ($\gamma_{00}$ is analogous to the intercept, $\gamma_{10}$ is analogous to the slope), but the error term is now the sum of three components. The terms in the parentheses (i.e., $\epsilon_{ij} + \zeta_{00i} + \zeta_{10i} * A_{uij}$) collectively represent error in this model. Because this is an important change with considerable implications, I discuss the contents of the error term in some detail, although nontechnically (see Singer & Willett, 2003, pp. 243–265, for a more complete discussion of the composite error term).

The residual term, $\epsilon_{ij}$, retains its usual meaning and assumptions. It is drawn from a normal distribution with a mean of 0, and its values are independent across persons and measurement occasions; that is, it is homoscedastic within and between participants. Each individual also has a constant added to each measurement occasion, the residual term $\zeta_{00i}$, and another residual, $\zeta_{10i}$, that interacts with $A_{ui}$. These latter two terms allow error to be correlated across measurement occasions, as might be expected when measurements are repeated for each individual. The product term, $\zeta_{10i} * A_{uij}$, allows error to be heteroscedastic within each individual. That is, the effect of this residual depends on the magnitude of $A_{uij}$. These latter two residual terms, $\zeta_{00i}$ and $\zeta_{10i}$, are also assumed to have means of 0 and to be drawn from a bivariate normal distribution, with unknown variances and covariance. These terms are homoscedastic across individuals (and constant across an individual’s measurement occasions).

To summarize, the models for the cross-sectional and longitudinal situations are different primarily in their error terms. The composite error term for the longitudinal case ($\epsilon_{ij} + \zeta_{00i} + \zeta_{10i} * A_{uij}$) allows error to be correlated and heteroscedastic across measurements. However, the heteroscedasticity comes only from the interaction term, $\zeta_{10i} * A_{uij}$, all other components are homoscedastic. Because we estimate all three residuals within the composite term when fitting this model, we can evaluate the consequences of nonlinear mapping for residuals that were originally homoscedastic. Specifically, we can evaluate whether any observed ordering relationship between items is the result of a nonlinear mapping between the underlying and observed variables by examining
whether $e_{ij}$ is systematically related to the predicted values of $B$.

To demonstrate this point, I created a longitudinal data set from the growth curve model just described, $B_{uij} = \gamma_{00} + \gamma_{10} * A_{uij} + (e_{ij} + \zeta_{0i} + \zeta_{1i} * A_{uij})$. I included five measurement occasions per individual; values of $A_{uij}$ increased as a linear function of time (i.e., measurement occasions) and ranged from 0 to 100. Consistent with the cross-sectional model, the population average for the intercept, $\gamma_{00}$, was 5, and the population average for the slope, $\gamma_{10}$, was 1.5. The three residuals, $e_{ij}, \zeta_{0i}, \zeta_{1i}$, were drawn from normal distributions with means of 0 and variances of 6, 1.5, and .05, respectively. (The variance of $\zeta_{1i}$ was set to a very small value to keep the model consistent with the simple synchrony relationship. Because it contributes to the composite error term as a product with $A_{uij}$ using a large value for the variance here would allow some of the cases to have zero or negative slope.) For simplicity, covariance between $\zeta_{0i}$ and $\zeta_{1i}$ was set at zero. I generated 200 cases (i.e., individuals) with this model, each with 5 measurement occasions, resulting in 1,000 (person-period) observations of $B_u$.

The top left panel of Figure 9 shows the scatter plot of values of $B_u$ as a function of $A_u$. The plot shows a strong linear relationship. However, it also shows the heteroscedastic properties of the composite error term; deviations from the best fitting line increase as $A_u$ increases. $A_u$ was used as predictor of $B_u$ in a mixed model; $\gamma_{00}$ and $\gamma_{10}$ were fixed effects, and $\zeta_{0i}$ and $\zeta_{1i}$ were random effects. The standard error covariance structure described earlier was assumed (Singer & Willett, 2003). It is not surprising that the model fit was good. The lower left panel of Figure 9 shows the residuals, $e_{ij}$, as a function of predicted values of $B_{uij}$; this pattern is homoscedastic.

Following the same logic developed for the cross-sectional case, I next created a nonlinear measure of $B_u$ by squaring it and allowing for measurement error: $B_{mij} = B_{uij}^2 + (e_{mij} + \zeta_{m0i} + \zeta_{m1i} * A_{uij})$. The latter two residual terms at the measurement level (indicated by the $m$ subscript) allow composite measurement error to be correlated and heteroscedastic across occasions. Analogous to the cross-sectional case, I assume that some unexplained differences are at the underlying level and others are at the measured level. Finally, I created the measured variable $A_{mij} = A_{uij} + e_{mij}$. Residual terms at the measurement level, $e_{mij}, e_{maj}, \zeta_{m0i}$, and $\zeta_{m1i}$ were independently drawn from normal distributions.

Figure 9. Consequences of nonlinear mapping for the observed data pattern and residuals: Longitudinal model. The figure shows the results for the longitudinal, growth curve model. The left-side panels show the results for the underlying variables, $A$ and $B$. The top left panel shows the relationship between items $A$ and $B$ at the underlying level. The lower left panel shows the pattern of residuals ($e_{ij}$), when $A_u$ is used to predict $B_u$, plotted as function of predicted values of $B_u$. The right-side panels show the results for the measures of $A$ and $B$. $A_m$ is a linear function of $A_u$; $B_m$ is a nonlinear function of $B_u$. The lower right panel shows the pattern of residuals ($e_{ij}$), when $A_m$ is used to predict $B_m$, plotted as function of predicted values of $B_m$. 
with means of 0 and variances of 6, 6, 1.5 and 1, respectively.

The upper right panel of Figure 9 shows the observed relationship between \( A_m \) and \( B_m \), which is now strongly curvilinear (I again linearly rescaled \( B_m \) to facilitate comparison). When \( A_m \) and \( A_m \) are used to predict \( B_m \) in the mixed model, the fit is still very good, but crucially for our discussion, the residuals are now related to the predicted values of \( B_m \). The lower right panel shows the values of the residual, \( \epsilon \), as a function of predicted values of \( B_m \). The cross-sectional case, the pattern of residuals is now heteroscedastic.

Nonlinear mapping between the underlying and observed variables has analogous consequences for cross-sectional and longitudinal models, and therefore can be used to test the alternative hypothesis that an observed developmental ordering relationship is due to a nonlinear mapping in both cases. Note that longitudinal models often explicitly allow error to be heteroscedastic within person, as the model used in the preceding example did by including the multiplicative term, \( \zeta_{ij} * A_{iij} \). However, even though the composite error term allowed for heteroscedasticity, the residual, \( \epsilon_{ij} \), which was homoscedastic at the underlying level, became strongly associated with the predicted values of \( B_m \) as a result of the nonlinear mapping.

Discussion

Research in developmental psychology has been bedeviled by an inability to test predictions about developmental ordering for continuously developing variables. Developmental ordering is a fundamental prediction about the developing system and a crucial form of evidence for evaluating developmental hypotheses. A promising test of developmental order, the shape of the relationship between two developing items, requires either (a) the assumption that one has interval-scale measures, an assumption that is difficult to justify and arduous to evaluate, or (b) evidence that the mappings between the measures and the underlying variables could not have produced the observed pattern of data.

I showed that a systematic, nonlinear mapping between an underlying variable and its measure has predictable consequences for the second moment. That is, specific types of nonlinear transformations leave different signature patterns in the residuals. Because the alternative hypotheses that are capable of explaining an observed data pattern (e.g., synchrony) can do so only through an easily definable set of transformations, examining the residuals for evidence of those transformations provides a method for disconfirming the alternatives.

Although this may initially sound complicated, the consequences for researchers are straightforward. The take-home points can be summarized as follows. First, it is important to keep in mind that measures are not direct reflections of the underlying variables of interest. This point is worth emphasizing if only because it is easy to lose sight of in the heat of data analysis. Second, the shape of the observed data pattern can be used to test developmental ordering hypotheses, but the patterns of residuals must also be examined. Third, a simple pair of rules for examining the residuals follows from the analysis presented here.

A Rule for Interpreting the Curvilinear Priority Pattern

When a researcher observes the classic, curvilinear developmental priority pattern, he or she can reject the competing alternative hypotheses (opposite partial priority and synchrony) if two conditions hold: (a) the predicted values of the item that appears to be developing more slowly are not positively related to the absolute values of the residuals for that item, and (b) the predicted values of the item that appears to be developing more rapidly are not negatively related to the absolute value of the residuals for that item.

A Rule for Interpreting the Linear Synchrony Pattern

When the linear, synchrony pattern is observed, the researcher can reject the competing alternative hypotheses (partial priority) if the following condition holds. Neither item has its predicted values negatively or positively related to the absolute values of its residuals.

Failure to observe the predicted heteroscedastic patterns places considerable pressure on the alternative hypotheses. The alternative hypotheses require specific nonlinear mappings, and those nonlinear mappings should produce specific heteroscedastic patterns. Therefore, applying these simple rules will allow researchers to eliminate alternative explanations of the observed data pattern. It is worth mentioning explicitly that patterns of heteroscedasticity other than those previously described are possible. These patterns, like homoscedasticity, are strong evidence against the alternative hypotheses (see Tabachnick and Fidell, 2001, for a discussion of alternative regression
methodologies for analyzing data with heteroscedastic residuals).

Limitations of Correlational Data and Other Caveats

Assessing the developmental relationship between two variables usually involves the interpretation of correlational data. Therefore, the usual caveats that accompany correlational designs apply. Conclusions from correlational data are always subject to the assumption that other unobserved variables are not driving the relationship. The approach described here, like other correlational methods, requires this assumption.

One consequence of this assumption is that if other covariates are theoretically important or are known to contribute to the development of the variables of interest, they may need to be included in the model. For example, if both language ability and executive function are believed to be important precursors for the development of children’s theory of mind (Carlson & Moses, 2001), both these variables should be included in the model. Failure to include the relevant variables can lead to misestimation of coefficients and standard errors and, most important for the current discussion, heteroscedasticity. Developmental researchers often address this problem by including global variables such as age, grade, measures of verbal ability, social and economic status, and so forth in the model. Although it is desirable to include measures of the specific variables involved in the developmental process, given the multiplicity of influences in development, using global variables has considerable practical appeal.

An additional caveat is also worth noting. If the homoscedasticity assumption is strongly violated at the underlying level or at the measurement level (i.e., error is truly related to the magnitude of the variables), or both, interpretive problems can result. Perhaps most seriously, true heteroscedasticity (i.e., a real association between error and the variable) might compensate for the heteroscedasticity induced by a nonlinear mapping. For example, error might be negatively related to the underlying variable, but an accelerating mapping function between the variable and measure would create a positive relationship. These opposing relationships might, in effect, cancel each other out. In this case, one might mistakenly conclude that homoscedasticity held and, therefore, erroneously interpret the observed ordering pattern. Interpreting the patterns of the residuals is no different from other correlational methods in this respect: Inferences drawn from correlational data always require the assumption that other variables are not distorting the observed relationship.

Developmental Order and Causal Modeling of Longitudinal Data

A wide variety of sophisticated methods for modeling causal relationships in longitudinal data have been developed in recent years (e.g., see Collins & Horn, 1991; Collins & Sayer, 2001; Singer & Willett, 2003). In general, these methods require interval-level measurement for continuous variables, reasonably large samples (how large depends on several factors, such as the complexity of the model), and three or more measurement occasions. One great strength of these models is that they allow for the use of lagged predictors. That is, they allow one to use variance in past performance to predict variance in performance at subsequent time points. Clearly, this provides a potentially powerful tool for researchers investigating developmental processes.

Although a review of these methods is beyond the scope of the current presentation, I briefly discuss one method, the latent difference score (LDS) model (McArdle & Hamagami, 2001), which has some important commonalities with the current approach and which may give a sense of how these rich models can be used to address developmental ordering hypotheses. The LDS model allows one to predict changes (i.e., differences) in scores across time, for example, the gain in peer status from Time 1 to Time 2. Note that a difference score computed across adjacent times (or estimated within a set of structural equations) is a simple linear approximation of rate of change during that period; it gives a straight-line approximation of the curve during that temporal interval. In one version of LDS, the dual change score model, change scores are predicted based on the weighted sum of the previous value of the variable and an individually varying constant. For example, one might model change in peer status as the weighted sum of previous peer status and an individually varying constant. This model can be extended to include additional predictors of change, including a second variable that is undergoing development at the same time. For example, peer status and social skill might undergo development during the same period, raising questions about their developmental ordering.

A more complex version of LDS, the bivariate dual change score model, is capable of addressing developmental ordering hypotheses. This model represents change in each of the two variables as the
weighted sum of its own value at the previous measurement occasion and the value of the other variable at that previous occasion (and an individually varying constant slope). Therefore, the model consists of two related equations; change in each variable is represented by one equation. The equations are related in that the outcome of one equation (i.e., change in the variable) feeds into the other equation (i.e., by changing the value of a predictor) at the next time point. These simultaneous equations can be estimated as a structural equation model. Hypotheses about developmental ordering can be evaluated by restricting the values of the model parameters. For example, a researcher could test whether children’s level of social skill predicted change in peer status by comparing the fit of the parameter. For example, a researcher could test whether children’s level of social skill predicted change in peer status by comparing the fit of the model with the social skill parameter included versus set to zero.

The LDS model and the approach presented in the current article both emphasize the form of the relationship between variables. The current approach does so by testing for the predicted functional form of the relationship and the alternative explanations for that relationship. The LDS approach does so by modeling the local slopes (i.e., for each temporal interval) using lagged predictors. For researchers with fairly large, longitudinal data sets, approaches such as the LDS model can test a wide variety of developmental ordering hypotheses, including hypotheses about reciprocal developmental influences. However, this power comes at a price—as McArdle and Hagamami (2001) pointed out, this method requires interval-scale data. Researchers interested in employing the LDS model might find it useful to test for systematic nonlinearity using the logic outlined here. A task for future work is to develop formal methods for linking the approach developed here with rich models such as LDS.

**Developmental Order, Measurement, and Types of Evidence**

An important issue in developmental research is what constitutes evidence for or against a developmental hypothesis. Clearly, there is no single answer to this question. Developmentalists marshal a wide variety of evidentiary types to test their hypotheses. However, I argue that the field has been unable to capitalize on one of its most fundamental types of evidence, developmental order, because tests of developmental order have not been available unless one is willing to take the position that development is saltatory (or that one’s continuous measures are interval scales). It is worth noting that if development is continuous, but a researcher’s measures are discrete, interpreting developmental ordering patterns becomes problematic (see Dixon & Moore, 2000, for a discussion). Therefore, converting one’s continuous measures into categorical measures, or collecting categorical measures of continuously developing variables, hides the problem but does not solve it. Researchers have been stuck in the difficult position of having theories that predict specific developmental orderings but no means to test adequately those predictions.

One might hope that the types of evidence used by new approaches such as computational modeling or developmental neuroscience might solve, or at least circumvent, the complexities discussed in this article. Although these are exciting areas that have made substantial contributions to our understanding of development, they too rely on the measurement of underlying variables and make predictions about developmental order. For example, in developmental neuroscience, Diamond (2000) proposed that the development of the cerebellum, as well as the prefrontal cortex, leads to functional gains in cognitive performance. Developmental changes in the cerebellum should be synchronous with changes in the relevant cognitive tasks. Johnson (2000) proposed that specialization of cortical processing occurs earlier for regions associated with output relative to regions associated with sensory processing. In both of these areas, researchers have worked hard to develop measures of their core constructs (e.g., cerebellar development, specialization of cortical processing). For example, Johnson measured cortical processing in infants using ERP, a measure of electrical changes that occur as collections of neurons fire in the cortex. The need to understand the nature of the mapping between the observed measure (e.g., ERPs) and the underlying variable (e.g., cortical processing) is not eliminated by the use of physiological measures. Whether measures are derived from techniques such as ERP, fMRI, and PET, or from behavioral assessments such response times, rating scales, and self-report, the relationship between the underlying variable and its measure has important implications for interpreting evidence of developmental ordering as well as other types of predictions (Busemeyer, 1980).

Computational models of developmental processes also rely heavily on developmental ordering. In general, computational models are either existence proofs, which demonstrate that a particular type of computational architecture can produce a phenomenon (e.g., learn a relationship), or ordering proofs, which demonstrate that a model goes through the
same developmental order as does the developing person. Both types of models can also offer new insights into the process under study and make new predictions. Consider the following examples of computational models as ordering proofs. Thelen, Schöner, Scheier, and Smith (2000) showed that a dynamic field model of infant reaching could reproduce the developmental changes observed in perseverative reaching, including the A-not-B error. They proposed that within the model a single, continuously developing parameter, analogous to the ability to keep relevant representations active in memory, explained the observed developmental order, as well as effects of various experimental manipulations.

Similarly, Cohen et al. (2002) presented a connectionist model of the development of causal understanding in infants. Their model showed the same developmental ordering observed in infants as they were learning to represent causal launching events (see also Buckingham & Shultz, 2000; Munakata, McClelland, Johnson, & Siegler, 1997). Computational models of developmental processes rely heavily on the evidence of ordering generated by traditional empirical studies.

The point here is that the field is not likely to escape from the joint complexities of developmental ordering and measurement any time in the near future, if at all. I use developmental neuroscience and computational modeling as examples, not to be critical of these valuable areas but because they employ advanced methods at a variety of levels and, therefore, might be considered on the cutting edge of the field.

Regardless of whether one works in neuroscience, computational modeling, or some other domain, developmental ordering is fundamental to developmental science because developmental theories are linked to data largely through the orderings they predict. The inability to test these predictions empirically breaks this important link. The approach presented here offers one way to test developmental order, thereby reconnecting developmental theory and with this crucial type of evidence.

**References**


