Infant and Child Development
Inf. Child Dev. **16**: 7–33 (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/icd.506

Wobbles, Humps and Sudden Jumps: A Case Study of Continuity, Discontinuity and Variability in Early Language Development

Marijn van Dijk[a,*] and Paul van Geert[b]
[a] *Open University of the Netherlands, The Netherlands*
[b] *University of Groningen, The Netherlands*

Current individual-based, process-oriented approaches (dynamic systems theory and the microgenetic perspective) have led to an increase of variability-centred studies in the literature. The aim of this article is to propose a technique that incorporates variability in the analysis of the *shape* of developmental change. This approach is illustrated by the analysis of time serial language data, in particular data on the development of preposition use, collected from four participants. Visual inspection suggests that the development of prepositions-in-contexts shows a characteristic pattern of two phases, corresponding with a discontinuity. Three criteria for testing such discontinuous phase-wise change in individual data are presented and applied to the data. These are: (1) the sub-pattern criterion, (2) the peak criterion and (3) the membership criterion. The analyses rely on bootstrap and resampling procedures based on various null hypotheses. The results show that there are some indications of discontinuity in all participants, although clear inter-individual differences have been found, depending on the criteria used. In the discussion we will address several fundamental issues concerning (dis)continuity and variability in individual-based, process-oriented data sets. Copyright © 2007 John Wiley & Sons, Ltd.

Key words: intra-individual variability; early language development; continuity; discontinuity; random sampling techniques; dynamic systems theory; microgenetic

One of the major achievements of individual-based, process-oriented approaches (dynamic systems theory and the microgenetic perspective) is the increase in the number of variability-centred studies that have appeared in the literature. The implication is that a growing number of researchers acknowledge the meaningfulness of intra-individual variability and show an interest in irregular aspects of

*Correspondence to: Marijn van Dijk, Open University of the Netherlands, The Netherlands. E-mail: Marijn.vanDijk@ou.nl

change. In line with these approaches, we have proposed to view developmental data as ranges of levels (see van Geert and van Dijk, 2002). These ranges are relevant, since they provide us with information on the flexibility and sensitivity of a system to its environment. Where the maximal value represents the maximal performance of a child under optimal circumstances, the minimal value indicates the vulnerability of this performance to other (less optimal) conditions. This flexibility and vulnerability might be different for different individuals or for the same individual at a different measurement occasion, i.e. for different phases in development. In this article we will take the argument a step further by proposing a technique that incorporates variability in the analysis of the shape of developmental change as described in microgenetic and dynamic systems studies.

Developmental change can literally take many shapes, ranging from the monotonous increase of an underlying continuous curve to a pattern of ranges or patches of variability. One important distinction among these different shapes of development concerns the difference between *continuity and discontinuity.* Discontinuities are of special interest because they may indicate increasing self-organization in the system, that is, a spontaneous organization from a lower level to a higher level of development (van Geert, Savelsbergh, & van der Maas, 1999).

In this paper, we argue for the incorporation of variability in the analysis of discontinuity. The focus will be on methodological aspects, more specifically the use of resampling techniques and various null-hypotheses of (dis)continuity. As we will demonstrate, results differ depending on the definitions and exact calculation methods employed. This article should therefore be seen as a general claim for a move towards an 'exploratory' type of statistics. Instead of holding on to a single analysis technique, we advocate the need to explore a series of statistical models that are tested with resampling procedures. Also, instead of perceiving continuity and discontinuity as a dichotomy, we argue that they might be conceived of as two extremes of one continuum. On this continuum, intermediate positions are possible, with the implication that a pattern might have some aspects of continuity and some aspects of discontinuity. Our final aim is to contribute to a better analysis and description of microgenetic data, including those collected within the framework of dynamic systems.

Continuity and Discontinuity

Although the meaning of the terms 'continuity' and 'discontinuity' differs across studies, a general definition of discontinuity is provided by van Geert *et al.*, (1999): 'we can define a discontinuity in the most general sense by identifying it with a stage transition. That is, if a stage is replaced by another stage, in the sense of consecutive sets of different equilibria, a discontinuity has occurred' [pp. XV]. In a discontinuity the growing variable jumps from one level (or stage) to the next without intermediary points. In practice, however, such a discontinuous curve is hard to distinguish from a (very) steep continuous curve. The reason for this is that the empirical data sampling is almost by definition discrete and thus easily misses points that lie along the steep increase. Thus, although the empirical representation of a developmental curve may show a discontinuity in the form of a lack of intermediary points during a sudden jump or increase, it is not automatically implied that the data refer to a discontinuous underlying process (see Figure 1). Figure 1(a) represents a continuous sigmoid curve of the form $y = 1/(\text{time}+\text{Exp}(-(t-0.5)/0.001))$. On a scale from 0 to 1, it shows a (seeming) discontinuity, however, a 'magnification' of the time scale (below) shows that the

DOI: 10.1002/icd

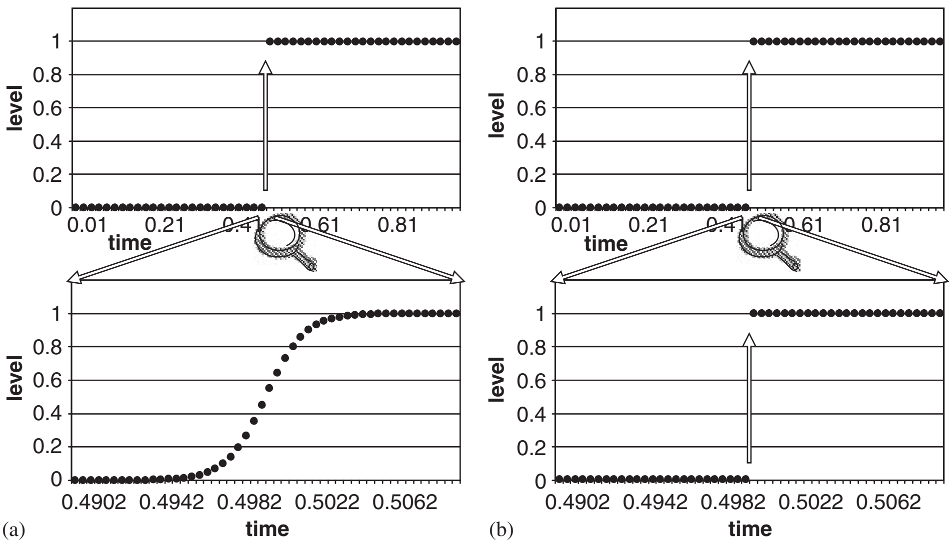

Figure 1. (a) Represents a continuous sigmoid curve of the form $y = 1$ / (time + Exp($-(t - 0.5)/0.001$)). On a scale from 0 to 1, it shows a (seeming) discontinuity; however, a 'magnification' of the time scale (below) shows that the function is continuous and has no gaps. (b) Consists of a step function $\{y = 0$ if $x < 0.5$; $y = 1$ if $x => 0.5\}$, which shows a gap on all scales of magnification and is thus discontinuous.

function is continuous and has no gaps. Figure 1(b) consists of a step function $\{y = 0$ if $x < 0.5$; $y = 1$ if $x = > 0.05\}$, which shows a gap on all scales of magnification and is thus discontinuous according to the mathematical 'epsilon–delta' definition. It should be noted that the terms 'spurt' and 'jump' are often used in relation to discontinuities, while in fact spurts and jumps in the data may refer to both continuous and discontinuous change.

For the types of curves that are of interest to developmentalists, the mathematical definition of continuity is very close to the intuitive description given above (Lakoff & Nuñez, 2000; Barile, 2006). The developmental functions at issue are all functions of time, for any point on the time axis, there is a value of the developmental function. For instance, at any hour, day, week, etc., a particular child is characterized by an average level of some psychological variable (this can be a score on a cognitive test, the number of complex language utterances recorded, etc), which evolves from 0 (or any convenient initial level) at the beginning of life and stabilizes at some value when the child has reached cognitive maturity. Let us assume the curve has a particular value at time t, namely L_t. The curve is continuous at time t, if for any moment arbitrarily close to time t, there is a value of the curve that is arbitrarily close to L_t. This approach to continuity, which is also known as the *epsilon–delta definition* (Barile, 2006), is very close to the notion of a gap. A continuous curve should not contain a gap. A curve that increases its value extremely rapidly is still continuous if there are no gaps in the curve (see Figure 1). In Figure 1(b), the last point of the lower part of the curve (at around time 0.5) is separated from the first part of the upper part by a gap. That is, for any point in time arbitrarily close to t, there is no value of the curve that is arbitrarily close to L_t.

A curve is a mathematical object that shows a gap only at the point of its eventual discontinuity. The question is, however, whether developmental data

DOI: 10.1002/icd

can be validly described in the form of curves that eventually show the type of discontinuity described by the gap criterion. To put it differently, are (underlying) curves indeed the descriptive format that developmentalists wish to extract from their data and, if they are not, is the standard gap definition of discontinuity still applicable to those data? Before answering this question, let us first take a closer look at an important property of developmental data, namely their variability within individuals.

Variability

In van Geert and van Dijk (2002), we have defined intra-individual variability as 'differences in the behaviour within the same individuals, at different points in time' (pp. 341). In quantitative developmental data, intra-individual variability is expressed as fluctuations between measurement points. Variability has largely been neglected until the introduction of dynamic systems and microgenetic perspectives. Both take a radical departure from the traditional approach to variability. This traditional view is voiced by a strong axiom in psychology called 'true score theory', which considers variability to be the result of measurement error (Cronbach, 1960; Lord & Novick, 1968; Nunnally, 1970). Dynamic systems theory and the microgenetic approach, on the other hand, claim that variability is developmentally meaningful and bears important information about the nature of the developmental process. For instance, Thelen and Smith (1994) explain that, in a self-organizing process, behavioural variability is an essential precursor of an attractor state (one preferred configuration out of many possible states). They state: 'Variability is revealed when systems are in transition, and when they undergo these shifts, the system is free to explore new and more adaptive associations and configurations' (pp. 145). Thelen and Smith encourage researchers to investigate the variability in their data. Inspired by this new definition, several studies have been performed that adopt Thelen and Smith's suggestion, they treat variability as data, and analyse it (for instance Bertenthal, 1999; Bertenthal and Clifton, 1998; de Weerth and van Geert, 1999, 2002; Newell and Corcos, 1993). Variability also features in the microgenetic approach. This focus on variability is shown, among others, in the work of Siegler (1994, 1996, 1997), Goldin-Meadow, Alibali, and Church (1993), Alibali (1999), Granott (1993), Fischer and Granott (1995), Fischer, Bullock, Rotenberg, and Raya (1993), Fischer and Yan (2002), Lautrey, Bonthoux, and Pacteau (1996), Lautrey (1993), Lautrey and Cibois (1994), Tunteler and Resing, (2002), Granott (1998, 2002), Pine, Lufkin, and Messer (2004) and Flynn, O'Malley, and Wood (2004).

The reason that we look at variable data is that we want to learn something about the causes of the observed phenomena. We agree with Miller (2002) that variability in cognitive performance can provide insight into the process, or 'generator', of change. The classical view is that the generator is a *dedicated internal* generator, with specific properties. For instance, in the case of developing intelligent behaviour, such a generator may be a 'device' or 'module' in the brain with a certain strength or 'power' that produces cognitive behaviour. In the classical view, the actual product (e.g. the score on a cognitive test) is caused by the internal generator (the 'device') plus a number of additional mechanisms or variables that randomly intervene with the working of the actual generator. Classical test theory aims at measuring the power of the cognitive device, and sees the additional variables as sources of measurement error. Dynamic systems theory, on the other hand, proposes the notion of a *distributed generator* that

DOI: 10.1002/icd

consists of all the factors or agents that causally contribute to the production of the observed variable at issue at a particular time and place. This generator participates in a developmental process. This implies that in our example, a score on a cognitive test is the product of the mutual and complex interaction between the capacities of the child and the context.[1]

The way a researcher looks at the nature of the data is closely related to the way he or she looks at the nature of the generator. The notion of an internal, dedicated generator mechanism implies that the observed data are the product of the generator plus a number of accidental influences. Thus, according to this belief, in order for the data to teach us something about the progress of the underlying generator, the errors must be disposed of, which in all likelihood will result in a smooth curve that corresponds with the assumed real level and change of the underlying generator. From a dynamic systems point of view, however, the generator is a complex process itself, the properties of which are represented, among others, in the way it fluctuates over time. Thus, the fluctuations are part of the information that the data can give us about the nature of the generator and therefore an analysis of these fluctuations is of central importance.

In the next section, we will present examples of microdevelopmental studies in which the issue of continuity versus discontinuity plays an important role and discuss the contribution of catastrophe theory.

Discontinuities in Developmental Studies

The question as to whether a developmental process is continuous or discontinuous is a relevant problem in (micro)developmental studies. Let us discuss an example from Fischer and Granott (1995). Their study consisted of the problem solving interaction of dyads learning to understand an unfamiliar device. One striking example of a particular dyad showed that there was an initial low level of understanding in both individuals, then a clear progression, which was characterized as '...non-linear, dynamic microdevelopment, with up and down oscillations, gradually moving to increasing complex interactions and representations' (p. 309). In short, the microdevelopmental pattern shows a change in the range across which the oscillations occur. It remains to be tested, however, whether this change is indeed gradual (continuous), or discontinuous.

As a second example, let us consider the balance scale tasks of Siegler and his colleagues (see Siegler & Chen, 2002). These balance scale tasks are often seen as a benchmark for the testing of the usefulness of various methodological and theoretical approaches in cognitive development. A recurrent finding of these studies is that a large majority of 5-year-olds consistently use rules to solve balance scale problems, while 3-year-old children rarely use rules systematically. Balance scale rules are ordered hierarchically. At first children use predictions only based on weight (Rule I), then only based on distance (Rule II), and finally they combine both quantitatively (Rule III). However, while rule use correlates with age, there exist obvious inter-individual and intra-individual differences. This means that variability (both inter-individual and intra-individual) is also present in the performance on balance scale tasks. Siegler and his colleagues employ the typical criterion for concluding that a child used a given rule, that 80% of the child's responses on a large number of items are predicted by that rule. The question remains how the frequencies of rule use develop. Is there a gradual increase in rule use across trials?

According to Janssen and van der Maas (2002), there is empirical evidence that at least one transition in the development of solving the balance scale

DOI: 10.1002/icd

problem—from rule I to rule II—is discontinuous. The theoretical framework for their conclusion is based on catastrophe theory (Thom, 1975), which can be considered a specific branch of dynamic systems theory. According to this theory, self-organizational processes can be classified into a limited number of characteristic patterns of discontinuous change, depending on the number of fundamental variables that determine the change. As such, catastrophe theory offers concrete models and criteria for discontinuities in developmental processes. All these catastrophes are discontinuities that are caused by changes in control variables (i.e. variables controlling the generator process that produces the changes in the measured variable). Of these seven catastrophes, the cusp is the simplest: it is controlled by two variables and there are two equilibrium states. The cusp can be fitted directly to data consisting of measurements for x, y and z (Cuspfit, Hartelman, 1996, based on the method of Cobb & Zachs, 1985). A second way of testing the cusp catastrophe model uses the qualitative characteristics of the cusp model. Catastrophe theory provides eight so-called catastrophe flags to test for the presence of a cusp model (Gilmore, 1981). Janssen and van der Maas (2002) have found evidence that the transition from rule I to rule II is discontinuous in that it is characterized by four (out of five testable[2]) catastrophe flags.

The method of catastrophe testing has several drawbacks. First, the direct fitting of a cusp catastrophe model is confined to data sets for which estimations of the two control variables are available and thus, by implication, is limited to phenomena that critically depend on only two control variables. In many domains of development, no real control parameters are available. However, the most serious problem with catastrophe testing is the fact that the underlying model does not acknowledge the variability as true meaningful data. Instead, the cusp catastrophe model is in fact based on a point source model, in the sense that the source of development (the generator) corresponds with certain 'true' levels of development, which can be represented as single values on a curve. Observed variability across the trajectory is considered to be noise added to this point source. Hence, the catastrophe model provides no answer to the question of how discontinuity can be demonstrated in a phenomenon that is in essence a change in its range properties.

Discontinuity and Ranges

The phenomenon we deal with concerns the development of a distributed generator, as above. An example, discussed in the preceding section, is the dyadic generator producing a changing understanding of an unfamiliar device. Another example is the generator that produces the verbal and non-verbal reactions of a child to a balance-scale question. If it is the essence of a phenomenon that it appears as a score range (due to variability), the model of this phenomenon should also be focused on that range, in such a way that the model must describe this range as accurately as possible. Therefore we define a discontinuity as, 'a transition from one variability pattern (range), to a different variability pattern, in the sense that these patterns are separated by a gap'.

Adopting an individual-based approach, we formulate the following assumptions. We may either have certain *assumptions about the underlying nature of this range* (for instance, that it can be described in terms of a normal—or some other—distribution), or we may assume that the best possible model of the range

DOI: 10.1002/icd

are the range data themselves (this assumption comes close to the basic assumption of the bootstrap method in statistics, see Efron and Tibshirani, 1993). In the latter case, we should use this range *without* further 'corrections' (for instance without the prior use of a smoothing procedure). In any case, the main question, which we introduced earlier, must now be answered: how can the gap-criterion of discontinuity (the epsilon–delta criterion of mathematics) be applied to discontinuities in ranges?

To begin with, time-serial data from developmental studies consist of sets of separate data points. That is, there are gaps everywhere. Does it make sense to apply the gap definition to something where gaps are everywhere? The trick is to consider a set of developmental, time-serial data as a single object, which in this case is the range. Continuity can be defined as set-membership, as belonging to the range. Any two data points within the range are, by definition, continuous. Any two data points, one of which lies in the range and one of which lies not in the range, are by definition, discontinuous. The set of observed data points is of course just a small subset of all the possible points that could lie within the range. Thus, how do we decide whether a particular, new data point, lies in the range or not?

To make this decision we envisage an imaginary magic bug that can jump from any point in the range to any other one, but cannot do anything else. Thus, the range is in fact defined by the set of possible jumps the magic bug can make, whatever point lies within the bug's radius of action is a potential member of the range. Let us proceed by adding a new data point to the series, in this case a new measurement of some developing variable. Does this new data point correspond to a continuity or a discontinuity? We can ask the bug. If he has a jump in his repertoire that can bring him from anywhere in the range to the new point or beyond, there is no gap from the bug's point of view between the range and the new point. Thus, the new point is continuous to the range. If he has no such jump in his repertoire, and thus cannot reach the new data point from somewhere in the range, there is a gap from his point of view, and the new data point is a discontinuity. The magic bug metaphor maps directly onto a mathematical definition of (dis)continuity in ranges defined by sets of data points. Central to the argument is the definition of a range of data points by the set of possible 'jumps' from one point to another within the range.

Figure 2 represents the temporal layout of a range across the time dimension. The new data point, at time 30, can have three values: (a), (b) or (c). Position (a) belongs to the range by definition. In order to determine the degree-of-membership of values (b) and (c), we need the range layout and the magic bug that defines the range of the set of the 'possible jumps'. Position (b) lies within the range of the bug's jumps, and its degree-of-membership is greater than 0 but smaller than 1. Position (c) cannot be reached by the jumping magic bug: there remains a gap between the bug's furthest position and position (c), (c) has a degree-of-membership of 0, and thus indicates a discontinuity at time 30.

In the remainder of this article we will demonstrate an approach to analyse (dis)continuity in developmental processes using the concept of developmental *ranges*. This approach is based on the exploratory use of resampling techniques, and shows the application of various statistical models of discontinuity to data of early language development. A further characteristic of discontinuity in developmental data is that it can take various forms, depending on the type of developmental process we are dealing with. Thus, the empirical criteria of discontinuity will vary with the type of process under study. However, they will all boil down to some form of anomaly that marks a point of transition. In the

DOI: 10.1002/icd

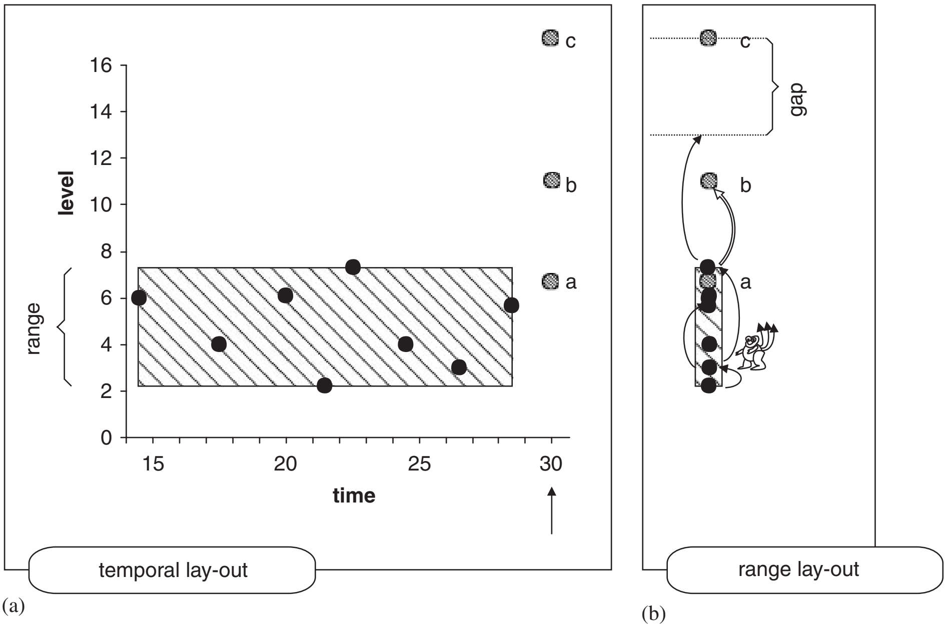

Figure 2. (a) The range covered by the hatched rectangle is continuous at time 30: the measurement at time 30 can be 'reached' from the range and there is no gap. (b) The hatched range is discontinuous at time 30: the measurement made at time 30 cannot be 'reached' from the range, hence, there is a gap at time 30 (for a formal definition of 'reachability', see text).

next section, we will introduce our case study. The discontinuity criteria applicable to this particular case will be presented in the method section.

Illustration: a Case Study on (dis)continuity in the Development of Early Prepositions

In this article, we will demonstrate the application of several criteria to test for discontinuities in a particular empirical example. This consists of the question of whether the acquisition of early *prepositions-in-contexts* shows discontinuous characteristics. Both issues (variability and discontinuity) are related to an important theme in the acquisition literature. Language development is often characterized in terms of stages or phases (for instance, in the case of Dutch, see Fikkert, 1998; van der Stelt & Koopmans-van Beinum, 2000; van Kampen & Wijnen, 2000). The question remains as to whether these stages are merely descriptive characterizations or whether they are true discrete stages (a sequence of periods in time with coherent elements that have a certain degree of stability), with discontinuous transitions.

In the language acquisition literature, there are only a handful of studies that explicitly report quantitative data on the developmental trajectory of prepositions. Interestingly, these suggest that this process might be stage-wise. For example, in his classic case-study Brown (1973) describes that initially, Eve omitted the prepositions 'in' and 'on' (in so-called 'obligatory' contexts) more than she supplied them. However, from the seventh measurement onwards, prepositions were constantly supplied in 90–100% of all obligatory contexts.

DOI: 10.1002/icd

Stenzel (1996) revealed a similar jump-wise pattern in a bilingual child. In the beginning, prepositions were only used very infrequently. He states, '[I]n the first phase, the absolute number of tokens is very low, and we find some strange distributions. In the second phase, the number of utterances containing prepositions is rather high in some recordings (and zero in others); and the pattern observed in the first phase vanishes.' [pp. 1036]. Thus, while the quantitative development of prepositions might be stage-wise, it is unknown whether the transition between these stages is discontinuous.

The issue of continuity versus discontinuity in the development of these prepositions-in-contexts can be approached from different theoretical standpoints. In the case of continuity there might be a gradual differentiation: from a restricted use to a more refined and flexible application of prepositions in a variety of distinct contexts. In this case, the child acquires the new prepositions context-by-context or type-by-type. From one conceptual standpoint, continuity can be expected if the order of acquisition follows the underlying conceptual complexity of prepositions (Clark, 1978; Johnston & Slobin, 1979), that is, if the acquisition of the related non-linguistic spatial knowledge is continuous. From the modular point of view, continuity can also be expected, as the result of the mapping of spatial concepts between three main cognitive modules, namely perception, action and language (van Geert, 1986). In the case of a discontinuous curve, the child suddenly discovers the way prepositions can be used to label spatial relations between objects. This discovery might function as a threshold, after which prepositions are used in abundance, while before their use having been seldom, and only in verb-like situations. A discontinuous trajectory is expected if, for instance, the required syntactic rules are acquired, and these rules generalize instantly across types and contexts. Here, the discontinuity stems from the sudden acquisition of a new structural category, namely the syntactical use of prepositions. This may lead to an instantaneous application of the preposition category to a wide collection of spatial contexts. However, while the question of continuity and discontinuity is certainly relevant, language has primarily been chosen as an exemplary data set to demonstrate the application of our methods. Our aim is to lay out a procedure for investigating variability and discontinuity in an integrated fashion. This procedure is applicable to a wide variety of developmental processes. In this sense, there is no difference between language development and other aspects of cognitive development such as cognitive skills (e.g. Fischer & Yan, 2002), strategy change in solving mathematical problems (e.g. Alibali, 1999), or scores on a balance scale test (e.g. Jansen & van der Maas, 2002).

METHODS

The Data set

In the present case study, four participants (Heleen, Jessica, Berend and Lisa) were followed over the course of a year from around age 1 year and 6 months to age 2 years and 6 months. At the beginning of the study, the infants were predominantly at the one-word stage, while at the end of the observation period their language showed various characteristics of the differentiation stage. At this stage, children generally acquire, besides the major lexical categories (such as nouns, verbs adjectives/adverbs), the word classes that have a primary syntactical function (such as articles, pronouns, prepositions). Also, children learn to use morphological and syntactical rules and, as a result, their sentences

DOI: 10.1002/icd

Table 1. Participant characteristics

Name	Ages	Number of Observations	Gender
Heleen	1;6,4–2;5,20	55	Female
Jessica	1;7,12–2;6,18	52	Female
Berend	1;7,14–2;7,13	50	Male
Lisa	1;4,12–2;4.12	48	Female

become longer and more complex. As an example, in the differentiation stage children, start using verb inflections (for a detailed description of the characteristics of the Dutch differentiation stage, see Gillis & Schaerlaekens, 2000). The participants were raised in a monolingual Dutch environment, with no apparent dialect. The children's general cognitive development was tested with the Bayley Developmental Scales 2/30 (van der Meulen & Smrkovsky, 1983) a few months before their second birthday. The scores were average to above average. All participants came from middle class families, who lived in suburban neighbourhoods in average to large cities in the Netherlands. For details on the participants, see Table 1.

The study is based on videotaped observations of spontaneous speech under naturalistic circumstances (the child's home). (For further details on the data collection and measurement design see van Geert & van Dijk, 2002). For the analysis of the speech samples, the child's utterances that contained a preposition were transcribed orthographically according to Childes conventions (MacWhinney, 1991). This was done by an experienced transcriber (the first author or intensively trained graduate students, one student per infant). Inter-observer reliability was calculated as the positive overlap ratio on the basis of an utterance-by-utterance comparison of a series transcripts out of a random sub-sample of the total data set. The resulting positive overlap ratios were 0.84 (Heleen), 0.88 (Jessica), 0.87 (Berend) and 0.87 (Lisa), which is considered to be adequate.[3] All prepositions that belong to the set of spatial prepositions were selected, but only if the context was spatial. The total set of spatial prepositions consisted of 'in', 'uit', 'op', 'af', 'voor', 'achter', 'tussen', 'over', 'bij', 'naar', 'onder', 'boven', 'binnen', 'buiten', 'door' (approximate translations are in, out, on, off, before/in front of, after/behind, between, over, near (to)/at, to/by, to, under, above, in/inside, out/outside, through). We counted the total frequency of prepositions that were uttered in a particular spatial context. All *distinct spatial contexts* were counted, excluding exact repetitions.

The Analyses

The analyses are based on individual trajectories, and consist of bootstrap and resampling procedures (for general discussions of these methods, see for instance Good, 1999, Manly, 1997; the procedures were carried out in Microsoft Excel, by means of a statistical add-in, Poptools, Hood, 2001). The question we address is: do the individual trajectories show discontinuities in the developing range?

Criteria of Discontinuity

The concept of discontinuity we adopted in this study is 'a transition from one variability pattern (range), to a different variability pattern, in the sense that these patterns are separated by a gap', a definition that assigns a central position to

DOI: 10.1002/icd

variability. Therefore, our testing criteria are designed to be sensitive to changes in these variability patterns. The question we intend to answer by means of these procedures is, to what extent is a continuous model capable of producing the ranges of our four participants. This question can be reformulated as, to what extent do we need a discontinuous model to produce the data of our four participants? The smaller the probability that a continuous model (plus random variability) reproduces our data, the less likely it is that the observed differences in our four participants are an accidental outcome of an undivided, continuous developmental trajectory.[4]

The Null-hypotheses

The null-hypotheses are based on various continuous models. These are, respectively, a (1) *linear* and (2) *quadratic* model on the one hand and non-linear models with symmetric noise based on (3) *Loess smoothing* or (4) *moving average smoothing* of the data on the other hand. Each of these models follows the trend in the data in a continuous fashion. In line with the continuity-hypothesis, variability is considered as noise. Consequently, for each of these models, a noise component was estimated by fitting a regression model to the residuals (the differences between the observed values and the values estimated by the continuous curves). The regression model specifies the expected variance of the noise for each point in time, under the assumption that the noise follows a normal distribution. A continuous null hypothesis model can thus be simulated by adding a Gaussian noise component to any point of the estimated continuous curve. We will proceed as follows, first, the continuous models will be estimated on the basis of characteristics of the observed data. Then, these models are used to simulate data sets. Each of these simulated data sets will therefore, by definition, be produced by the continuous model. If these simulated models are capable of producing the statistical indicators of discontinuity that we observed in our participants, the null-hypothesis of underlying continuous development cannot be rejected. It is also possible to perform a meta-(resampling)analysis to this data, but since our approach is individual-based, our emphasis is on the individual curves.

The Criteria

In this study, we looked for three indicators that are likely to occur in any kind of discontinuity. The *first* is the existence of two clearly distinct sub-phases or sub-patterns. If a discontinuity occurs, it is likely to be evident in the form of two distinct sub-patterns, e.g. distinct patterns of data in terms of distinct score ranges. The question is, what is the probability that the sub-patterns that we think correspond with a discontinuity, in fact result from a continuous model with symmetrical noise. The *second* indicator is the existence of an anomaly in the data at the moment of the discontinuous shift. We expect that such an anomaly can be observed in the form of an unexpectedly large, local peak or 'spike' in the data. The peak results from the (presumed) sudden emergence of some new form (rules governing the use of spatial prepositions). In addition, there is a fair chance that new forms are likely to be abundantly used right after their initial discovery (a sort of novelty effect). Finally, the *third* indicator is based on the assumption that the discontinuity corresponds with the fact that the data consist of two discontinuous sets of scores, expressed in terms of degrees-of-membership to a set. Referred to as the gap criterion, this is defined by the analogy of the magic bug.

DOI: 10.1002/icd

The use of three different operationalizations of discontinuity and of four null hypothesis models is inspired by the method of converging evidence. The more the criteria converge on the finding of a discontinuity in an individual trajectory at the same point in time, the more credibility is given to the conclusion that the discontinuity is real. The choice of our exploratory procedures and the principle of convergent evidence do not necessarily imply that the proposed set of criteria and models is exhaustive or sufficient. Other developmental phenomena might require discontinuity criteria that are eventually better adapted to the particular nature of the phenomenon in question. Starting from the idea of converging evidence, we have taken the definitions of discontinuity that we have provided above as a reasonable starting point, assuming that their transformations into testable criteria provide us with a sufficient answer to the question whether or not our data show discontinuous changes.

Criterion 1: Significantly Different Sub-patterns

The definition of a discontinuity was formulated above stated, 'a transition from one variability pattern (range), to a different variability pattern, in the sense that these patterns are separated by a gap' (see Figures 1 and 2). The first and strictest definition of discontinuity requires the search for two distinct variability patterns and assumes a model of two clearly discernible phases that can be characterized as two distinct series. On the other hand, if the null-hypothesis of continuous development is true, the trajectories basically consist of a single undivided pattern. If this is the case, the observed frequencies with which prepositions occur will randomly fluctuate around a single, continuous trajectory. With such random fluctuation, however, it is not unlikely that some arbitrary divisions of the entire data set in two subsets will result in sub-patterns that also show some apparent discontinuity, as defined by the different-sub-pattern criterion. The critical issue is, what is the probability that the continuous models with random fluctuations produce a two-stage difference that is of the same magnitude as the two-stage difference found in our data? The smaller the probability, the less likely it is that the observed distances in our four subjects are an accidental outcome of an undivided and thus continuous developmental trajectory.

In various publications, Fischer has claimed that it is not possible to observe discontinuities in the developmental data if the data are based on what he calls the 'functional level', which means the level of unsupported performance (see Fischer & Rose, 1994; Fischer & Yan, 2002). In order to discover eventual discontinuities, we must monitor the child's optimal level of performance, which is achieved by giving the child support. The discontinuities in that optimal level, if any occur, are indicators of stage-wise changes in the child's performance. Although we cannot provide our participants with support, we can nevertheless obtain an estimation of their optimal preposition production by simply taking the local maximum of the preposition frequencies (the local maximum is the maximum value in a window of observations, in this case five consecutive observations).

Thus, by monitoring the change in local maximum levels, it is possible to demonstrate a discontinuity in the maximal values, in the use of spatial prepositions. However, we announced above to 'investigate whether a discontinuous transition from one variability pattern to a different variability pattern exists.' If, by variability pattern we mean the width of the range within which the use of spatial prepositions can vary from observation to observation,

DOI: 10.1002/icd

the focus on maximal values cannot, by definition, be used to show that such a discontinuity exists, simply because the maximal level alone is not an index of variability. Thus, we may think of a sudden shift in the distance between the extremes, as an indicator of the presence of two sub-stages.

Statistical Procedure

As a measure for the maximum performance level, we took the Progressive Maximum. This starts with the value of the first observation and then takes the maximum of a progressively widening window that expands from the second up to the last observation (see van Geert & van Dijk, 2002 for a description and justification). In the same manner, for the local minimum, the Regressive Minimum was calculated. The regressive minimum starts with the value of the last observation and then takes the minimum of a progressively widening window that expands from the last up to the first observation (see also van Geert & van Dijk, 2002). We now calculate the difference between the two extremes (the difference between maximal and minimal performance) and take this as our *key value*. We now divide the data series in two parts, taking all possible divisions (beginning with data point 1–5, compared to the rest of the data, i.e. 6 to last, then 1–6 against 7 to last, and so forth). For each subset, we select the maximal *key values* obtained. The *difference* between these two key values defines the maximal *distance* between the two sub-patterns for this particular data subset. The place where the distances differ maximally from each other marks the position of the assumed discontinuity (see Figure 3).

This procedure is carried out for the original data set and compared with a similar procedure for the data sets simulated on the basis of the continuous null hypothesis models. This whole procedure of generating a simulated series and calculating the maximal difference between the two subsets in the simulated data is repeated a great number of times (in this case 5000 times). Finally, we count the number of times the simulated continuous model has produced a maximal

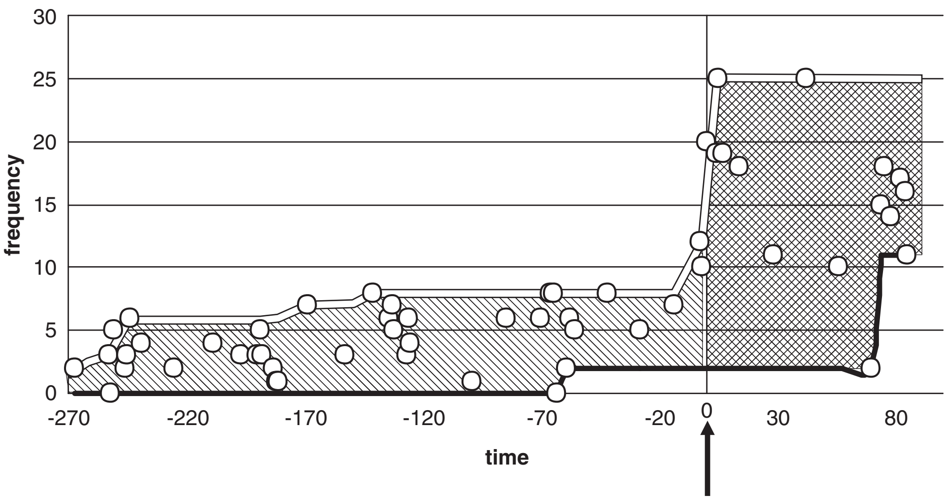

Figure 3. Data from Heleen show a potential discontinuity at time 0, based on the assumption that the range expands discontinuously at time 0.

difference between the subsets that is as big as or bigger than the maximal difference between subsets that we have obtained from our observed data series. This number, divided by the number of simulations, gives us the *p*-value of the observed difference under the null hypothesis as specified.

Criterion 2: The Peak Model

While Criterion 1 (see previous section) conceptualizes discontinuity in terms of two distinct variability patterns, the second criterion focuses on the aspect of the 'gap' at moment of the transition. This gap might be conceptualized as a 'sudden change' in the curve. Thus, with this criterion, we identify a discontinuous transition if the trajectory shows an *unexpectedly* large peak or spike, at some point in time. This peak might reveal *the moment* at which the system loses its stability and shifts into a different variability pattern. As mentioned earlier, there is a fair chance that the sudden emergence of a new skill or ability is characterized by a (probably short) time of abundant use of that new skill or ability. This transient performance peak, if any occurs, will be added to the sudden shift in the sub-patterns described in the previous section, and thus marks the underlying transition even more clearly.

Statistical Procedure

We conceptualize a peak as the *maximal positive*[5] *difference* between an expected and an observed value in the trajectory. With this criterion, we have defined a peak in two distinct ways: an *absolute* peak and a *relative* peak. In case of the *absolute peak*, variability at a specific point in time is defined as the absolute difference between the observed (or simulated) data point and the expected point, based on the underlying continuous model. This difference is the residual value, which is used to define the absolute peak. In the case of the *relative peak*, these residuals (for all points in time) are divided by the expected value, i.e. the value of the continuous model at the corresponding point in time. We have made the distinction between absolute and relative peaks because the degree of variability is often strongly related to the central tendency in a distribution. As an illustration, assume a trajectory that shows a simple increase from value 1 to value 50. In this example, a residual of 4 is considered to be more salient if it occurs around an observation with value 5 if when it occurs around value 45. While the absolute peak is strictest and signals unexpected spikes in the data, the relative peak criterion is sensitive to smaller peaks that might be meaningful in the context that they appear in.

In our statistical procedure, finally, we take the peak (the maximal value of either absolute or relative residual) in the data as our key value and test it against the peaks found on the basis of a great many (5000) simulated continuous models, with distribution characteristics similar to those of the data.

Criterion 3: Membership Procedure

Previously, we have postulated that if a discontinuity occurs, it shows itself in the form of two distinct sub-patterns. According to the epsilon–delta definition of (Barile, 2006), a curve is continuous at time t, if for any moment arbitrarily close to time t, there is a value of the curve that is arbitrarily close to L_t. On the other hand, a discontinuous curve contains a *gap* between the two stages or phases,

DOI: 10.1002/icd

however close they are to each other in time. We have shown that this epsilon–delta or gap criterion can be generalized to data sets or ranges, by means of the magic bug metaphor. Formally speaking, the bug's jumps define a set of data that are within reach of the bug. That is, all data points within reach of the bug are members of the set and have a set or range membership equal to 1. Any new data point that is within reach of the bug's repertoire of jumps are also members of the set, and those that are not within reach are not members of the set or range. The third criterion relates the existence of a gap to the concept of *set membership* that stems from *Fuzzy Logic* (McNeill & Freiberger, 1993).

Fuzzy logic rests on the idea that all things are characterized by the degree to which they are members of a class. In modern fuzzy logic, 'objects' (observations, objects, properties, etc.) are assigned a degree-of-membership to a particular category (see Ross, 1995, for a particularly clear technical introduction; see further Kosko, 1993, 1997; McNeill & Freiberger, 1993; Nguyen & Walker, 1997; von Altrock, 1995, provide a highly accessible introduction). In mutually exhaustive classes, objects have a degree-of-membership of either 1 or 0. For instance, a particular piece of furniture is either a chair or not a chair (for instance, it is a bench). In fuzzy logic, an object can have any degree of membership between 0 and 1. For instance, the piece of furniture is a chair that looks like a bench, e.g. has a degree of membership of 0.8 to the category 'chair'. Maximal ambiguity arises if an object has a degree of membership equal to 0.5. In this case, the object is 50% part of a category, and 50% not. According to the membership criterion, we conceive of the data set as either a single 'set', or as a series of 'sets' (in principle two). All data values can be assigned a degree-of-membership to each set. In the simplest case, the case of continuity, the data consist of a single set, and thus all values in the data set, have a degree-of-membership of 1 to that set. If the data are truly discontinuous, the data are cut into two consecutive sets if a new data point emerges that is not a member of the set of data points already present (say at time t). A data point belongs either to the first or to the second set, thus a data point has a degree-of-membership of 1 to the first set and 0 to the second set. Data points can also be somewhere in between and thus have a degree-of-membership of, for instance, 0.5 to the first and 0.5 to the second set.

Statistical Procedure

Set membership is easy to define. Suppose that we define data points a, b, c to m as members of a set S. Now we collect a new data point, n and wish to know if it also belongs to the set. Recall that the magic bug defined set membership by its possible jumps. A possible jump is for instance the jump from a to c, which is a jump of length $|a–c|$. Any member of the set is a possible starting point for the bug. Take for instance point f. Since all points that can be reached by the bug are members of the set, the point that lies at distance $f \pm |a–c|$ is also a member of the set. The degree-of-membership of (potential) members of the set is defined as follows. If f is the set's maximal value, if $|a–c|$ is the sets greatest distance and if n lies not within reach of $f \pm |a–c|$, the degree-of-membership of n to the set S is 0. More precisely, there exists a *gap* between n on the one hand and the set S on the other hand.

All points that have been assigned to the set (i.e. a to m) have, by definition, a degree-of-membership of 1. All points that lie between any two possible members of the set (e.g. a point lying between a and d) have, by implication, a degree-of-membership of 1. All other points have a degree-of-membership

DOI: 10.1002/icd

which is equal to the likelihood that such points are within reach of the bug (lying within reach refers to the membership by implication: remember that if a is a member and b is a member, a new point n is also a member if it lies between a and b). Thus, the degree-of-membership of n is defined by the sum of probabilities of all the possible 'jumps' for which n lies within reach. Since a jump is defined mathematically as $m_1 \pm |m_2–m_3|$ (for m_1, m_2 and m_3 arbitrarily chosen full members of the set), the degree-of-membership is defined by the sum of probabilities of such triplets. A practical way of calculating degrees-of-membership is by randomly sampling triplets of values from a data set S, calculate the resulting $m_1 \pm |m_2–m_3|$ value and repeating that a great many times. By doing so we can numerically approximate the likelihood of occurrence of any possible value and thus determine its degree-of-membership to the set.

The set S can be specified by taking an initial set (e.g. the first four to five data points, ranging from time t_1 to t_4), determine if the first data point outside the set (i.e. the point at time t_5) is a member, add it to the set if its degree-of-membership is satisfactory (e.g. > 0.8, > 0.95 or whatever seems reasonable) and continue until we meet a data point whose degree-of-membership is smaller than some pre-established value (in this case 0.05). This data point defines a discontinuity in the data.

RESULTS

Visual Inspection of the Developmental Curves

In order to get a general impression of the development of prepositions we plotted the trajectories of the four participants in Figure 4. On the basis of visual inspection, we aligned participants to their first major peak in their trajectories. Using this representation, all participants show their first major 'jump' at point

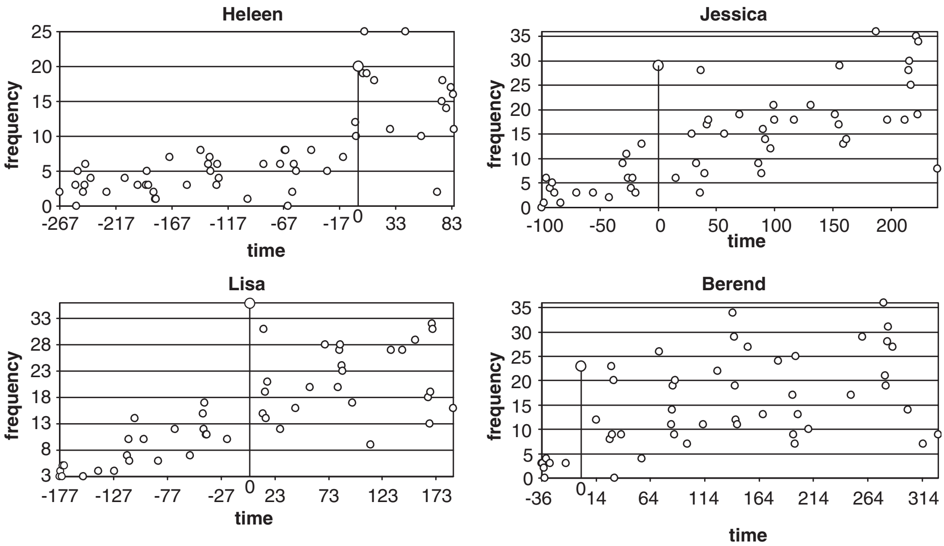

Figure 4. The individual trajectories of prepositions-in-context, Heleen, Jessica, Lisa and Berend.

DOI: 10.1002/icd

zero on the x-axis. The other numbers on the axis represent the number of days between the other observations and this point zero.

Visual inspection of the individual trajectories shows that most participants can be described in terms of two distinct phases. First, they show a period in which prepositions-in-context only occur occasionally, usually the phase before point zero. In this phase, prepositions fluctuate mildly between values lower than 10. As an exception, Lisa's initial values reach around 15. Then, further on, all trajectories are characterized by strong fluctuations, which occur after the first major outbursts of prepositions. Inter-individual differences are clearly visible. For instance, the timing of the transition between this first and second phase differs between the four participants. While Heleen shows a relatively long initial phase, Berend only shows seven values that might be characterized as such. It can be questioned whether Berend actually displays the initial phase we described above. However, since we have no data before measurement point −38, we can only speculate on the existence of a first phase as found in the other participants.

Thus, solely on the basis of visual inspection, trajectories of these four participants may be described in terms of two distinct phases. First, there is an initial phase in which there are only a restricted number of prepositions-in-contexts. Secondly, this phase is followed by a more advanced phase in which prepositions are used in many different contexts, but whose usage is highly variable, dependent on context.

Results of Criterion 1: Sub-pattern Criterion

Since the analyses were performed on the individual data, we will report the results for each participant separately (see Table 2 and Figure 5). As Table 2 shows, the discontinuity estimated by the procedure lies exactly at the estimated point 0 in our four samples. Put differently, the algorithmic method replicates our eyeball statistics with regard to the onset of the transition for all children. However, when we look at the p-values with regard to the different null-hypotheses, we see large differences between individuals. In the case of Berend, we see only one low p-value (for the linear model) and in Heleen and Jessica, there are no p-values below 0.05. Furthermore, in the case of Lisa, the p-values of the different null hypothesis models converge on the conclusion that the discontinuity is likely to be real, and thus unlikely to be caused by just accidental factors (see Table 2).

Table 2. p-values of the differences between the sub-patterns, based on 5000 simulations

	p-values				
	Estimated position	Linear model	Quadratic model	Symmetric residuals model Loess	Symmetric residuals model moving average
Berend	0	0.04	0.22	0.14	0.13
Heleen	0	0.16	0.1	0.12	0.11
Jessica	0	0.07	0.1	0.07	0.09
Lisa	0	0.03	0.06	0.001	0.01

DOI: 10.1002/icd

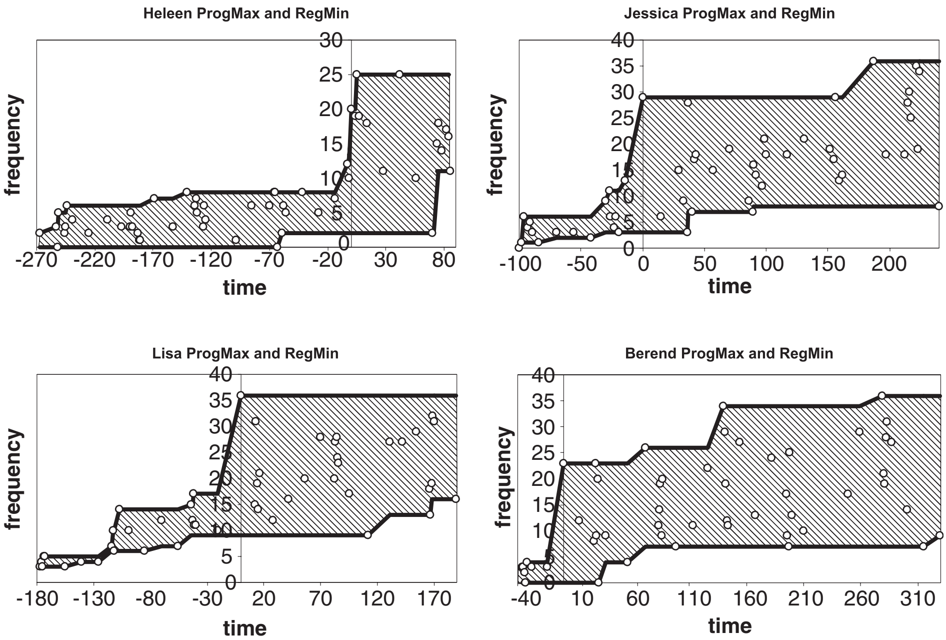

Figure 5. Moving progmax-regmin graph of prepositions-in-context, Heleen, Jessica, Lisa and Berend.

Table 3. *p*-values of the peak method based on 5000 simulations

	p-values				
	Estimated position	Linear model	Quadratic model	Symmetric residuals model Loess	Symmetric residuals model moving average
Relative peaks					
Berend	0	0.53	0.43	0.96	0.96
Heleen	5	0.39	0.35	0.99	0.98
Jessica	0	0.01	0.03	0.36	0.53
Lisa	0	0.004	0.04	0.09	0.36
Absolute peaks					
Berend	0	0.68	0.31	0.14	0.13
Heleen	75	0.1	0.32	0.09	0.08
Jessica	0	0.16	0.16	0.06	0.06
Lisa	0	0.23	0.17	0.01	0.05

Results on Criterion 2: The Peak Criterion

Table 3 shows that most of the positions of the discontinuity estimated with both the relative and the absolute peak methods correspond with those based on visual inspection (with the exception of Heleen, where the calculated position is close to the estimated in the case of the relative peak, but far off in the case of the absolute peak). More importantly, the analysis reveals that the peak criterion is

DOI: 10.1002/icd

less robust or consistent than the sub-pattern criterion. Using the relative peak as a criterion, peaks greater than expected on the basis of the continuous model occur only in two children, Jessica and Lisa, and only for the linear and quadratic continuous models. Using the absolute peak as a criterion, greater than expected peaks occur with Lisa ($p < 0.05$) and to a certain degree Jessica and Heleen ($0.05 < p < 0.10$), but only for the continuous models based on symmetric residuals.

Results Obtained on the Basis of the Membership Procedure

In Figure 6, we have plotted the degree-of-membership with the initial data set for each participant. As we have argued before, if the degree-of-membership drops dramatically (to very low values), there is a strong indication for a discontinuity at that point in time. If, on the other hand, membership remains high (and does not drop beneath 0.5), or drops gradually, the values are likely to belong to the initial set and there is no discontinuity. Gradual drops imply that the ranges are gradually expanding, which may be expected on the basis of continuous change.

Figure 6 shows that there is a sharp drop in the degree-of-membership for all participants. For instance, in the case of Heleen (Figure 6, upper left), the degree-of-membership drops from 1 to 0.02 at point 0 (the moment of the first jump according to visual inspection). The degree-of-membership expresses the probability that a particular data point can be reached from within a specified range of data, and thus resembles a p-value to a certain extent. Thus, purely on the basis of visual inspection of Figure 6, there is a strong indication for a discontinuity at time point 0 in the case of Heleen, as with the other participants. It is interesting to observe that in all participants, the degree-of-membership

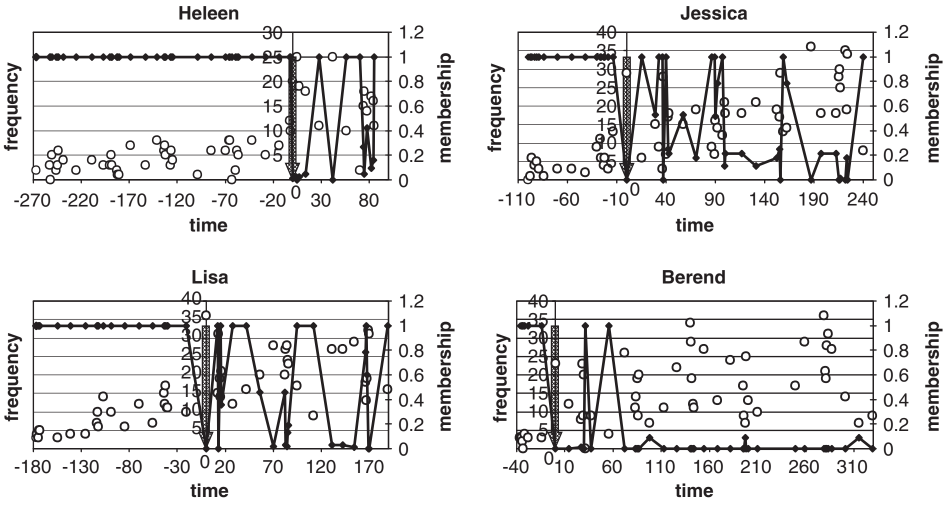

Figure 6. Data of Heleen, Jessica, Lisa and Berend. Data up to time 0 define the hypothesized initial range. The membership function (y-axis to the right) suddenly drops to (about) 0 for the four data sets and thus demonstrate that time 0 corresponds with a discontinuity in terms of range-membership.

DOI: 10.1002/icd

Table 4. *p*-values of the membership criterion

Membership	*p*-values			
	Linear model	Quadratic model	Symmetric residuals model Loess	Symmetric residuals model moving average
Berend	0.07	0.2	n.a.*	n.a.
Heleen	0.08	0.09	0.09	0.07
Jessica	0.04	0.04	0.08	0.08
Lisa	0.06	0.04	0.07	0.09

Note: For Berend, the significance of the membership criterion using the Loess model and the moving average model could not be assigned because of the early peak in his data (too few data points in the first stage of the trajectory).

fluctuates dramatically between 0 and 1 *after* point 0. This means that after the values that are discontinuous, values occur that *are* continuous with the initial set. This points in the direction of a combined model (continuous *and* discontinuous) in which certain values belong, and others do not belong, to the initial set. The existence of a fluctuation between an original level or range and a new, more advanced level or range is comparable to the flag of 'bimodality' in the catastrophe model. To test the likelihood of a sharp drop in membership against the null-hypotheses we have discussed earlier, we must run similar randomization tests as in the preceding sections. The result of these tests is presented in Table 4, which shows three *p*-values below 0.05 (two for Jessica and one for Lisa). However, most other *p*-values are around 0.06–0.09. Although this is not significant in the classical sense, this is an indication that we cannot be certain of a continuity either. It simply indicates that in most cases, the chance that this result occurs is around 6–9% (in each individual case).

Although we have argued for the need to focus on individual trajectories, the results we have found applying the membership criterion might be very suited for a meta-analysis. In most other cases presented in this article, a meta-analysis over all participants does not add much information, because the results are highly sensitive to a single (significant) outlier. However, in this case, we observe a fairly homogenous picture of *p*-values below 0.10. A meta-analysis can estimate the probability that an underlying continuous model produced this pattern across individuals. Since the *p*-values for the membership criterion simply express the probability that a continuous null-hypothesis model shows the membership drop observed in the data, the meta-analysis is easy to calculate. The four data sets show the membership drop and thus, the score of sudden membership drops is 4 out of 4. The probability that four series based on a null hypothesis model is simply the product of the probabilities that each of the models shows the sudden drop. These products in all models are (much) smaller than 0.001.

DISCUSSION

Continuity and Discontinuity in the Development of Prepositions

On the basis of visual inspection we hypothesized that the development of prepositions-in-contexts of these four participants shows a characteristic pattern of two phases. First, there is an initial phase in which prepositions are only used

DOI: 10.1002/icd

Table 5. Summary of the results, number of tests that yielded p-values <0.05 or exact estimated positions

	Sub-pattern criterion (5 tests)	Peak criterion relative (5 tests)	Peak criterion absolute (5 tests)	Membership criterion (4 tests)	Total set (19 tests)
Berend	2	1	1	0	**4**
Heleen	1	0	0	0	**1**
Jessica	1	3	1	2	**7**
Lisa	4	3	3	1	**11**

in a restricted number of contexts, followed by a second phase in which prepositions are used in many different contexts and are highly variable. In order to test whether the transition between these assumed phases represent a discontinuity three criteria were used, with various null hypothesis models.

Table 5 provides a general overview of the results of all testing procedures (frequency of low p-values). For simplicity, we have taken a cut-off point of <0.05, comparable with what is common in statistics. We have added this table to provide an overview, but realize that it displays a very crude representation of the data (see Tables 2–4, for the exact p-values and estimated positions). It might be argued that in some cases a cut-off point of 0.05 is somewhat arbitrary and may not be optimally adequate. In the case of the membership criterion, for instance, we have shown that most p-values were between 0.04 and 0.09. Thus, there is a probability of 4–9% that the continuous models replicate the data of *each individual* of *each specific criterion*. In this particular case, where the four individual cases yield approximately similar results, a meta-analysis over the group data might provide an idea of how probable it is that the discontinuity indicator in question can be found in each individual case, if the underlying processes are in fact continuous. We have seen that for the membership criterion this probability is far below 0.001.

With regard to the results of the individual pathways, Table 5 shows that the evidence for a discontinuity is strongest for Jessica and Lisa, and weaker for Berend. In the case of Heleen, the evidence was especially weak, only one out of all tests resulted in a p-value of below 0.05, which is at (or even below) chance level. This conclusion seems somewhat peculiar in the case of Heleen, since purely on the basis of visual inspection, her data show the clearest stepwise pattern. However, for data that show such a stepwise pattern, it is difficult to distinguish between a (very) rapid continuous increase in the variable at issue, and a discontinuity. This result stresses the fact that simply by using eye-ball statistics, a very steep continuous curve can seem to be 'discontinuous', while, in fact, the data can be (re)produced by a underlying continuous model. The case for discontinuity is especially strong in participant Lisa. Not only do the majority of the tests result in significant p-values, the p-values in 5 other cases are low as well (0.09 and below). This implies that the different null hypothesis models converge on the conclusion that the discontinuity is likely to be real, i.e. highly unlikely to be caused by just accidental factors.

Fundamental Considerations

The question remains as to which, and how many, of the discontinuity criteria must be met in order to decide that a change pattern is indeed discontinuous. The

DOI: 10.1002/icd

testing for discontinuities from a catastrophe theory's framework has encountered exactly the same problem. In principle, only the presence of all testable catastrophe flags indicates a cusp catastrophe. The examples of 'catastrophe flag studies' (for instance Ruhland & van Geert, 1998; van der Maas, Raijmakers, Hartelman, & Molenaar, 1999; Wimmers, 1996) show that this strict criterion is seldom (if ever) met. This leads us to the question of how many flags and additional indicators (such as the model fitting and other observations) are necessary and sufficient to indicate a discontinuity. The use of three different criteria of discontinuity and of four null hypothesis models is inspired by the method of converging evidence. If all criteria and hypotheses converge on the finding of a discontinuity in an individual trajectory at the same point in time, the conclusion that the discontinuity is real obtains more credibility. However, we can still question how many criteria must be met, and in what portion of the participants they must be observed, in order to speak of a discontinuity.

We have seen that the study of the development of prepositions-in-context shows clear inter-individual differences. Not only does the timing of the first major increase differ (for instance Berend has his first major increase very early on in his trajectory while Heleen shows a relatively long first phase) but more interestingly, the shape of the transitions differs across participants. This is also reflected in the individual results of the tested criteria and various null-hypothesis models. For instance, Lisa shows 'the most discontinuous' developmental trajectory, while Heleen turned out to be 'most continuous', in spite of the fact that her data provide the clearest sign of a stepwise pattern. The finding of these inter-individual differences is hardly surprising (see Shore, 1995, for an overview). In fact, in the domain of language development inter-individual differences are well documented (see Beers, 1995; Bates, Dale, & Thal, 1995, for an overview). With regard to the fundamental issue of whether the acquisition of prepositions consists of a gradual differentiation of prepositions or a sudden discovery of the prepositional phrase, we cannot provide a conclusive answer. As it turns out, there are clear indications of discontinuity in some infants and less clear (or no) indications in others. This finding might suggest that in the acquisition of prepositions, both processes, namely gradual acquisition and sudden jumps, are present at the same time and might alternate.

There are two important fundamental assumptions about discontinuities that are relevant here. The first, that we mentioned earlier, is that continuity and discontinuity are usually treated as categorically distinct and that a developmental process is thus considered to be either of the one or the other form. The second issue is that when testing for discontinuities, the default null hypothesis is assumed to be continuity. Only if continuity can be rejected, can the developmental process be regarded as discontinuous. The catastrophe flag studies we discussed (Ruhland, 1998; van der Maas *et al.*, 1999; Wimmers, 1996) also take this position. However, it should be noted that the null-hypothesis might also be reversed and reformulated as, 'a process involving a transition between two states (e.g. no prepositions versus use of prepositions) is discontinuous unless proven otherwise'. This position is at least equally defendable since an underlying discontinuous process is compatible with the fact that in a considerable number of individuals the process is not overtly distinguishable from a continuous process.

As an alternative to the traditional approach to continuity and discontinuity, we propose to conceive of discontinuity as a collection of characteristics that each describe different aspects of the discontinuity. In this case 'pure' continuity would be described as a complete absence of these characteristics, while 'pure'

DOI: 10.1002/icd

discontinuity indicates the presence of all characteristics. These two situations can be described as the two extreme positions on a continuum, between which many intermediate positions are possible. Characterizing a developmental process by describing which aspects of it are continuous and which aspects are discontinuous may provide more information than merely describing this same process in terms of the traditional either/or-definition of discontinuity. This is especially the case regarding the fact that the finding of one of these extremes (namely the discontinuity-extreme) is quite unlikely. Thus, a discontinuity model can explain why certain children show signs of discontinuity and others do not. Our new definition provides room for additional criteria to define discontinuity, since in most cases the precise number of control parameters is undefined. It should also be noted that, in terms of observable variables, a continuous model can be a special case of a discontinuous one, whereas the reverse does not hold. The sudden emergence of a new behavioural mode (a new skill, grammatical structure, etc.) that characterizes a discontinuity does not imply that the old behavioural mode suddenly disappears. In fact, the two modes co-exist for a while (a situation which is clearly demonstrated by the fold in the cusp model). Eventually, the replacement of the old mode by the new may occur in a linear, continuous fashion. In that case, the frequency of the new mode increases gradually. It is likely that children will differ in the form of the transition between the old and the new mode, resulting in patterns that are clearly discontinuous in some children and considerably more continuous in others. A comparable point of view has been defended by Sternberg and Okaggaki (1989). They state that intellectual development is not either continuous or discontinuous, but that it is simultaneously continuous and discontinuous with respect to different dimensions of development. They propose that instead of asking the question 'is development continuous or discontinuous?' we should ask 'what are the sources of continuity and discontinuity in intellectual development?'.

Another advantage of this approach is that the null-hypothesis can be formulated in various ways, dependent on the discontinuity criteria in question. Where some criteria may have continuity as the null-hypothesis, others may hold discontinuity as the starting point. Since the sufficiency question is no longer central in the analysis, this may lead to a much more subtle characterization of the developmental process under study. Thus, in our view, continuity and discontinuity are distinct categories, but there is a range where they overlap and where the developmental phenomena have properties that are partly continuous, and partly discontinuous.

We have emphasized before that this list of criteria is not exhaustive, but that it is also more than just an arbitrary choice. Criteria for discontinuity should try to account as much as possible for the peculiar properties of the developmental phenomenon under study. Our example, the use of spatial prepositions, shows its development primarily through its frequency of occurrence, in different contexts and in different grammatical constructions. What we observe in early language acquisition is the emergence of a new linguistic element, starting out with initially a very low frequency (restricted use of certain categories in only certain contexts), to a flexible and more grammatical use later on. In light of these properties, it is defendable to use the criteria put forward in the present article as primary indicators of discontinuity. However, further analysis of the developmental trajectories (both with regard to curve fitting or modelling and in combination with qualitative analysis) is needed in order to come to a better understanding of how continuity and discontinuity are related in the acquisition of spatial prepositions.

DOI: 10.1002/icd

Assuming an individual-based approach, a basic assumption of microgenetic studies, the inter-individual differences we have found in this study should be taken as seriously as intra-individual variability. Since the level of the developmental variable under investigation is a result of the dynamic interaction between the developing child and the interaction with the environment, causes of individual differences can be located in all control parameters that drive the process of development. To conclude, this implies that only if all participants show continuous trajectories can a continuous population model be considered sufficient to describe development. If the trajectories of some children show indications of discontinuity, a bifurcation might be assumed, parallel to catastrophe theory. In the case of such a bifurcation, there are two extreme positions, strict continuity versus strict discontinuity, but there are many positions in-between that do better justice to the complex and dynamic nature of developmental processes.

Notes

1. This assumption is akin to the idea of 'soft assembly' (Thelen & Smith, 1994).
2. It is important to note that not all flags can be observed in longitudinal data. These are the test for hysteresis, divergence, divergence from linear response and critical slowing down. See for a description Gilmore (1981).
3. See for a discussion of the positive overlap ratio van Geert and van Dijk (2003).
4. Technical details of the procedures can be obtained by contacting the authors.
5. Because of the nature of our developmental data (low initial values and a general increase), we only use positive differences. Conceptually, it is also possible to search for negative outliers as an indication of a discontinuous decrease, that is, if the data allows for this.

REFERENCES

Alibali, M. (1999). How children change their minds: strategy change can be gradual or abrupt. *Developmental Psychology, 35*(1), 127–145.

Barile, M. (2006). 'Epsilon-Delta Definition.' From MathWorld—A Wolfram Web Resource. Retrieved September 2006 from http://mathword.com/Epsilon-DeltaDefinition.html

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 96–151). Oxford: Basil Blackwell.

Beers, M. (1995). *The phonology of normally developing and language-impaired children*. Dissertation, University of Amsterdam.

Bertenthal, B. (1999). Variation and selection in the development of perception and action. In G. Savelsbergh, H. van der Maas & P. van Geert (Eds.), *Non-linear developmental processes*, Vol. 175 (pp. 105–121). Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Bertenthal, B., & Clifton, R. (1998). Perception and action. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology,* Vol. 2. *Cognition perception and language* (pp. 51–102). New York: Wiley.

Brown, R. (1973). *A first language: The early stages*. London: Allen & Unwin.

Clark, E. V. (1978). Strategies for communicating. *Child Development, 49*, 943–959.

Cobb, L., & Zachs, L. (1985) Applications of catastrophe theory for statistical modeling in the bio-sciences. *Journal of the American Statistical Association, 80*(392), 793–802.

Cronbach, L. J. (1960). *The essentials of psychological testing*. New York: Harper & Brothers.

DOI: 10.1002/icd

De Weerth, C., & van Geert, P. (1999). Transitions and variability in infant behaviour. In G. Savelsbergh, H. van der Maas, & P. van Geert (Eds.), *Non-linear developmental processes*, Vol. 175 (pp. 3–14). Amsterdam: Royal Netherlands Academy of Arts and Sciences.

De Weerth, C., & van Geert, P. (2002). Changing patterns of infant behaviour and mother–infant interaction; intra- and interindividual variability. *Infant Behavior and Development, 24*, 347–371.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Fikkert, P. (1998). The acquisition of Dutch phonology. In: S. Gillis & A. de Houwer (Eds.). *The acquisition of Dutch* (pp. 163–222). Amsterdam: Benjamins.

Fischer, K. W., Bullock, D., Rotenberg, E. J., & Raya, P. (1993). The dynamics of competence: How context contributes directly to skill. In R. H. Wozniak & K. W. Fischer (Eds.), *Development in context: Acting and thinking in specific environments* (pp. 93–117). Hillsdale, NJ: Erlbaum.

Fischer, K. W., & Granott, N. (1995). Beyond one-dimensional change: parallel, concurrent, socially distributed processes in learning and development. *Human Development, 38*, 302–314.

Fischer, K. W., & Rose, S. P. (1994). Dynamic development of coordination of components in brain and behavior: A framework for theory and research. In G. Dawson & K. W. Fischer (Eds.), *Human behavior and the developing brain* (pp. 3–66). New York: Guilford.

Fischer, K. W., & Yan, Z. (2002). Darwin's construction of the theory of evolution: Microdevelopment and explanations of variation and change in species. In N. Granott & J. Parziale (Eds)., *Microdevelopment: Transition processes in learning and development*. (pp. 294–318). Cambridge: Cambridge University Press.

Flynn, E., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science*, *7*(1), 103–115.

Gillis, S., & Schaerlaekens, A. (Eds.). (2000). *Kindertaalverwerving, een handboek voor het Nederlands* [*Child Language Acquisition, a handbook for Dutch*]. Groningen: Martinus Nijhoff Uitgevers.

Gilmore, R. (1981). *Catastrophe theory for scientists and engineers*. New York: Wiley.

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review, 100*, 279–297.

Good, P. I. (1999). *Resampling methods: A practical guide to data analysis*. Boston: Birkhauser.

Granott, N. (1993). Patterns of interaction in the co-construction of knowledge: Separate minds, joints efforts and weird creatures. In R. H. Wozniak & K. W. Fischer (Eds.), *Development in context: Acting and thinking in specific environments* (pp. 183–207). Hillsdale, NJ: Erlbaum.

Granott, N. (1998). Unit of analysis in transit: From the individual's knowledge to the ensemble process. *Mind*, *5*(1), 42–66.

Granott, N. (2002). How microdevelopment creates macrodevelopment: Reiterated sequences, backward transitions, and the zone of current development. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 213–242). New York, NY, US: Cambridge University Press.

Hartelman, P. (1996). *Cuspfit* [Computer software], University of Amsterdam.

Hood, G. (2001). *Poptools* [Computer software]. Pest Animal Control Co-operative research Center (CSIRO), Canberra, Australia.

Jansen, B. R. J., & van der Maas, H. L. J. (2002). Evidence for the phase transition from rule I to rule II on the balance scale task. *Developmental Review, 21*(4), 450–494.

Johnston, J. R., & Slobin, D. I. (1979). The development of locative expressions in English, Italian, Serbo-Croatian and Turkish. *Journal of Child Language, 6*, 529–545.

Kosko, B. (1993). *Fuzzy thinking: The new science of fuzzy logic*. New York: Hyperion.

Kosko, B. (1997). *Fuzzy engineering*. Upper Saddle River, NJ: Prentice Hall.

Lakoff, G., & Nuñez, R. E. (2000). *Where mathematics comes from. How the embodied mind brings mathematics into being*. New York: Basic Books.

Lautrey, J. (1993). Structure and variability: A plea for a pluralistic approach to cognitive development. In R. Case & W. Edelstein (Eds.), *The new structuralism in cognitive development: Theory and research on individual pathways* (pp. 101–114). Basel, Switzerland: Karger.

DOI: 10.1002/icd

Lautrey, J., Bonthoux, F., & Pacteau, C. (1996). Le traitement holistique peut-il guider le traitement analytique dans la categorization de visages?/Can holistic processing guide analytic processing for face categorization? *Annee Psychologique, 96*, 225–254.

Lautrey, J., & Cibois, P. (1994). Application of correspondence analysis to a longitudinal study of cognitive development. In D. Magnusson & L. R. Bergman (Eds.), *Problems and methods in longitudinal research: Stability and change* (pp. 190–211). Cambridge, England UK: Cambridge University Press.

Lord F. M., & Novick, R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

MacWhinney, B. (1991). *The Childes project, tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Manly, B. F. (1997). *Randomization, Bootstrap and Monte Carlo methods in Biology* (2nd ed.). Boca Raton: Chapman & Hall.

McNeill, D., & Freiberger, P. (1993). *Fuzzy logic*. New York: Simon & Schuster.

Miller, P. H. (2002). Order in variability, variability in order; why it matters for theories of development. *Human Development, 45*, 161–166.

Newell, K., & Corcos, D. M. (1993). *Variability and motor control*. Champaign, IL: Human Kinetics.

Nguyen, H. T., & Walker, E. A. (1997). *A first course in fuzzy logic*. Boca Raton: CRC Press.

Nunnally, J. C. (1970). *Introduction to psychological measurement*. New York: McGraw-Hill.

Pine, K. J., Lufkin, N., & Messer, D. (2004). More gestures than answers: children learning about balance. *Developmental Psychology*, 1059–1067.

Ross, T. J. (1995). *Fuzzy logic with engineering applications*. New York: McGraw-Hill.

Ruhland, R. (1998). *Going the distance; a non-linear approach to change in language development*. Dissertation, University of Groningen.

Ruhland, R., & van Geert, P. (1998). Jumping into syntax: Transitions in the development of closed class words. *British Journal of Developmental Psychology, 16*, 65–95.

Shore, C. M. (1995). *Individual differences in language development*. Thousand Oaks: Sage Publications.

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science, 3*, 1–5.

Siegler, R. S. (1996). *Emerging Minds. The process of change in children's thinking*. New York: Oxford University Press.

Siegler, R. S. (1997). Concepts and methods for studying cognitive change. In E. Amsel & K. A. Renninger (Eds.). *Change and development: Issues of theory, method, and application* (pp. 77–97). Mahwah, NJ: Erlbaum.

Siegler, R. S., & Chen, Z. (2002). Development of rules and strategies: Balancing the old and the new. *Journal of Experimental Child Psychology, 81*, 446–457.

Stenzel, A. (1996). Development of prepositional case in a bilingual child. *Linguistics, 34*, 1029–1058.

Sternberg, R., & Okaggaki, L. (1989). Continuity and discontinuity in intellectual development are not a matter of 'either-or'. *Human Development, 32*, 158–166.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development op cognition and action*. Cambridge, MA: MIT-Press.

Thom, R. (1975). *Structural stability and morphogenesis: an outline of a general theory of models*. Reading, MA: Benjamin.

Tunteler, E., & Resing, W. C. M. (2002). Spontaneous analogical transfer in 4-year-olds: A microgenetic study. *Journal of Experimental Child Psychology, 83*, 149–166.

van der Maas, H., Raijmakers, M., Hartelman, P., & Molenaar, P. (1999). Reaction time recovery after perturbation. In G. Savelsbergh, H. van der Maas & P. van Geert (Eds.), *Non-linear developmental processes*, Vol. 175 (pp. 159–168). Amsterdam: Royal Netherlands Academy of Arts and Sciences.

van der Meulen, B. F., & Smrkovsky, M. (1983). *BOS 2-30. Bayley ontwikkelingsschalen, handleiding*, Bayley Developmental Scales, Manual. Lisse: Swets & Zeitlinger.

van der Stelt, J., & Koopmans-van Beinum, F. (2000). Klankperceptie en klankproductie in het eerste levensjaar [sound perception and sound production in the first year of life]. In S. Gillis & A. Schaerlaekens (Eds.), *Kindertaalverwerving, een handboek voor het Nederlands* [*Child language acquisition, a handbook for Dutch*] (pp. 93–185). Groningen: Martinus Nijhoff Uitgevers.

DOI: 10.1002/icd

van Geert, P. (1986). In, on and under: An essay on the modularity of infant spatial competence. *First Language*, *6*, 7–28.
van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget Vygotsky and beyond. *Psychological Review*, *5*, 634–677.
van Geert, P., & van Dijk, M. (2002). Focus on variability, new tools to study intra-individual variability in developmental data. *Infant Behaviour and Development, 25*, 340–374.
van Geert, P., & van Dijk, M. (2003). Ambiguity in Child Language, the problem of inter-observer reliability in ambiguous observation data. *First Language, 23*(3), 259–284.
van Geert, P., Savelsbergh, G., & van der Maas, H. (1999). Transitions and non-linear dynamics in developmental psychology. In G. Savelsbergh, H. van der Maas & P. van Geert (Eds.), *Non-linear developmental processes*. Vol. 175 (pp. XI–XX). Amsterdam: Royal Netherlands Academy of Arts and Sciences.
van Kampen, J., & Wijnen, F. (2000). Grammaticale ontwikkeling [grammatical development]. In S. Gillis & A. Schaerlaekens (Eds.), *Kindertaalverwerving, een handboek voor het Nederlands* [*Child language acquisition, a handbook for Dutch*] (pp. 225–286). Groningen: Martinus Nijhoff Uitgevers.
von Altrock, C. (1995). *Fuzzy logic and neurofuzzy applications explained.* Upper Saddle River, NJ: Prentice-Hall.
Wimmers, R. (1996). *Grasping developmental change: theory methodology and data*. Academic Dissertation, University of Groningen, NL.

DOI: 10.1002/icd