Using Developmental Trajectories to Understand Developmental Disorders

Michael S. C. Thomas
Birkbeck College, London

Dagmara Annaz
University College London

Daniel Ansari
University of Western Ontario, Canada

Gaia Scerif
University of Oxford, United Kingdom

Chris Jarrold
University of Bristol, United Kingdom

Annette Karmiloff-Smith
Birkbeck College

Purpose: In this article, the authors present a tutorial on the use of developmental trajectories for studying language and cognitive impairments in developmental disorders and compare this method with the use of matching.
Method: The authors assess the strengths, limitations, and practical implications of each method. The contrast between the methodologies is highlighted using the example of developmental delay and the criteria used to distinguish delay from atypical development.
Results: The authors argue for the utility of the trajectory approach, using illustrations from studies investigating language and cognitive impairments in individuals with Williams syndrome, Down syndrome, and autism spectrum disorder.
Conclusion: Two conclusions were reached: (a) An understanding of the underlying mechanism will be furthered by the richer descriptive vocabulary provided by the trajectories approach (e.g., in distinguishing different types of delay) and (b) an optimal design for studying developmental disorders is to combine initial cross-sectional designs with longitudinal follow-up.

KEY WORDS: development, trajectories, delay, deviance, disorders

When researchers investigate behavioral deficits in individuals with developmental disorders, a common methodology is to proceed as follows: The disorder group is matched with two separate typically developing control groups, one match based on chronological age (CA) and a second match based on mental age (MA) derived from a relevant standardized test. If the disorder group shows an impairment compared with the CA-matched group but not with the MA-matched group, individuals with the disorder are considered to exhibit developmental delay on this ability. If, by contrast, the disorder group shows an impairment compared with both control groups, then the disorder group is considered to exhibit developmental deviance or atypicality (see, e.g., Hodapp, Burack, & Zigler, 1990; Leonard, 1998).

Recently, an alternative methodology has been increasingly applied to the study of disorders based on the idea of developmental trajectories or growth models (Annaz, Karmiloff-Smith, & Thomas, 2008; Jarrold & Brock, 2004; Karmiloff-Smith, 1998; Karmiloff-Smith et al., 2004; Rice, 2004; Rice, Warren, & Betz, 2005; Singer Harris, Bellugi, Bates, Jones, & Rossen, 1997; Thomas et al., 2001, 2006). In this alternative approach, the aim is to construct a function linking performance with age on a specific experimental task and then to assess whether this function differs between the typically developing group and the disorder group.

In this article, we describe the trajectory approach in more detail. In particular, we focus on a method that collects data at a single point of

measurement for typically developing and disorder groups and constructs cross-sectional developmental trajectories. We compare this method with the matching approach, bringing together in a single place many of the concepts and issues raised by the use of trajectories. To anchor our discussion, we contrast the matching and trajectory methodologies in the context of the idea of developmental delay.

The concept of delay is of interest because it is widely used in the study of developmental disorders to classify children's cognitive abilities. However, it has several shortcomings. As a mechanistic explanation of slower development, delay can amount to little more than a re-description of behavioral data indicating that the disorder group has produced scores and error patterns similar to those of younger typically developing controls—with no elaboration of the causal mechanisms by which this similarity may have arisen. For example, if delay were a causal mechanism, one might imagine that two straightforward predictions would follow: (a) assuming delay serves to modulate the rate of development in the cognitive system, performance in the disorder group should eventually reach the same endpoint as in the typical population; (b) as a parsimonious mechanistic explanation, delay should be the same across all cognitive domains. In many cases, neither pattern is observed in those individuals who are described as having developmental delay.

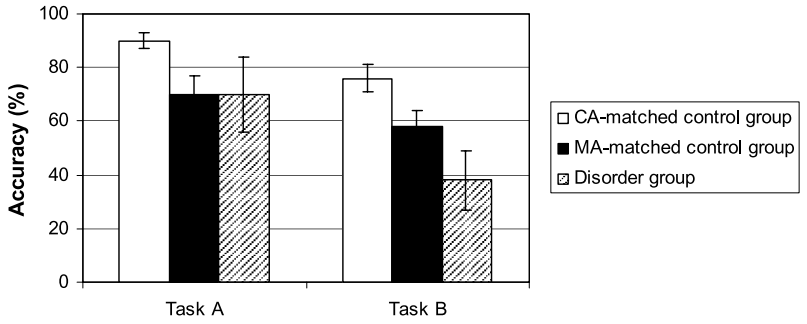
In the current context, our particular interest is in the use of delay as a descriptive term. In the following sections, we will illustrate the utility of the developmental trajectories approach by demonstrating how "delay," in fact, encompasses several different behavioral patterns that may ultimately require different mechanistic explanations. We begin by reviewing the traditional matching methodology used in the empirical investigation of disorders such as developmental dyslexia, specific language impairment, autism, Down syndrome, Williams syndrome (WS), Velo-Cardio-Facial syndrome, Turner syndrome, and Fragile X syndrome. We then discuss the developmental trajectory approach and show how it delineates different forms of delay. In two further sections, we illustrate the use of trajectories with a number of examples drawn from our own studies and then consider practical issues that arise in their use, such as interpreting null findings and validating cross-sectional trajectories via longitudinal follow-up. Finally, we consider how the matching and trajectory approaches allow us to decide whether a given pattern of development can be classified as qualitatively atypical (deviant, disrupted) rather than delayed—a distinction that many have argued is key in the study of developmental impairments of language and cognition.

Methodology 1: Individual or Group Matching

The use of CA-matched and MA-matched control groups to study developmental deficits has its origins in a theoretical debate on intellectual disabilities that contrasts the developmental and difference stances (e.g., Bennett-Gates & Zigler, 1998; see Hodapp & Zigler, 1990, for discussion of the debate in the context of Down syndrome). Difference theorists view learning disability as caused by underlying organic dysfunction, producing specific deficits in cognitive functioning and qualitatively atypical cognitive development. By contrast, developmental theorists view this characterization as only applying to a subset of individuals; additionally, there will be a group of individuals with learning disability who fall at the extreme lower end of the distribution of normal individual variation. These individuals will show the same overall qualitative pattern of development as nonimpaired individuals, including a similar sequence of developmental milestones and a similar structure to their intelligence. Although, by definition, one would expect the disorder group to exhibit impairments compared with CA-matched controls, an extreme-normal-variation group should look indistinguishable from a group that is individually matched on an MA measure that indexes the stage of developmental progression. The developmental and difference stances are echoed in the distinction between delay and deviance found in the literature on specific language impairment (SLI). Several decades of research in that field have attempted to establish whether the linguistic characteristics of children with SLI resemble those of younger typically developing children or are qualitatively different from anything seen in typical development (see Leonard, 1998, for a review and discussion).

The development and difference positions identify developmental processes in different individuals. However, the dichotomy is often applied to different component cognitive abilities within the same individual. For example, Figure 1 depicts the type of data that is often reported using the matching method (usually analyzed using *t* tests, analyses of variance [ANOVAs], or chi-square [χ^2] tests). In the example shown, performance is contrasted on two tasks to assess whether a developmental dissociation is present, perhaps to test a theory that the abilities tapped by the two tasks develop independently. In Figure 1, the disorder group performs at a lower level than the CA-matched group on both tasks. On Task A, the disorder group performs in line with MA-matched controls, whereas on Task B there is a deficit compared with MA-matched controls. A common interpretation would be that the disorder group is impaired/atypical/deviant on Task B, whereas

Figure 1. Example of data from a matching design. CA = chronological age; MA = mental age.



on Task A they are delayed rather than impaired. Where the experimental tasks tap areas of weakness in a disorder, individuals with the disorder are clearly expected to perform below the level of CA controls, and so for convenience, this latter control group is sometimes omitted (see, e.g., Clahsen & Almazan, 1998; van der Lely & Ullman, 2001).

There are two ways in which control groups can be matched to the disorder group. One can seek to carry out individual matching, where for each individual in the disorder group, a typically developing individual is selected with the same CA or MA; or one can be content that the mean CA or MA of the entire control group matches the mean CA or MA of the entire disorder group. Group matching is less desirable if the distribution of ages or abilities differs between control and disorder groups, as spurious differences in behavior could arise from this disparity. Alternatively, individual matching inserts a selection requirement that may reduce the generalizability of the findings (Mervis & Robinson, 2003). Group matching is less demanding on recruitment and may be adopted for practical reasons. Hereafter, we will combine these two methods and refer to them jointly as the *matching approach*.

Designs with MA-matched control groups rely on the use of standardized tests to match the level of developmental progression in the disorder group. This necessarily means that the group comparison is theory dependent: It is important for experimenters to be aware that they are taking a theory-driven view on what standardized test adequately measures developmental progression in the domain that the experimental task is thought to tap (from the range of standardized tests available; see Yule, 1978). For example, in tasks exploring disorders of language development, the experimenter might match the MA group according to a standardized test of receptive vocabulary, or productive vocabulary, or receptive grammar. In a typical receptive vocabulary test, the individual has to point to one of four pictures that corresponds to the word they have heard. However, it is a theoretical assumption that performance on such a standardized test is the correct single measure of the domain to assess developmental progress for, say, an experimental task exploring semantic priming in visual word recognition. One alternative to relying on any single measure of MA is to use composite MA measures. These measures average across a set of standardized tests to produce a "verbal" MA or even a "global" MA. However, frequently the point of investigating a given disorder is that performance is unequal across cognitive domains or even within domains (e.g., within language, between vocabulary and grammar). By contrast, the control group will tend to have more closely correlated abilities on all the subtests. The result of composite MA measures can be the selection of a control group that exceeds the ability of the disorder group on some standardized measures but falls short on others, compromising the interpretation of any task differences (Jarrold & Brock, 2004; cf. Klein & Mervis, 1999). The choice to select an MA group according to a composite measure is another theoretically driven decision made by the experimenter.

Once a theory-driven decision has been made about an appropriate MA group and once the data have been collected, there is a sense in which the experimenter is committed to this theoretical position. There is little flexibility to employ alternative measures of MA. One response to this constraint is to recruit multiple MA-matched control groups using different measures of MA, one per theory about which standardized test is relevant, with an attendant increase in the size and cost of the experiment. This approach may generate multiple conclusions about delay and deviance, if some MA-matched groups are equivalent in their performance to the disorder group whereas others are ahead or fall behind.

In practical terms, the matching method needs to avoid floor effects or ceiling effects on the task measures and standardized tests, as these render interpretation of results difficult or impossible (Strauss, 2001). For example, if a participant is at floor, his or her real ability level is unmeasured because we do not know how far

below floor the ability level falls—the measure is no longer working. Preferably, the CA, MA, and disorder groups should all be in the sensitive range of the tests and, at the very least, the MA and disorder groups should be in the sensitive range. This may limit the matching technique in cases where individuals with disorders have severe deficits, because there may be no age-equivalent performance in the typically developing population. The matching methodology is optimal when the disorder group covers a very narrow age range and/or when the experimental measure is only sensitive around a particular age. It is less advantageous when groups are averaged over a wide age range, which can sometimes be the case in studies of rare developmental disorders. In such cases, the group mean performance may mask a fairly wide range of individual performance, again limiting interpretability and inference to causal mechanism.

Finally, MA matching relies on the use of age-equivalent scores from standardized tests. For a given test score, one can derive the age at which the average child from the (typically developing) standardizing population achieved the same score. Some caution is required when using test-age–equivalent scores, as they are silent on the variability present in the standardizing population at each age (McCauley & Swisher, 1984). Many of the typically developing children may have scored some distance below (or above) the average age-equivalent score in the standardization sample, yet disparities of this nature are not treated as deficits (or hyperfunctioning) as they sometimes are in disorder groups.

Methodology 2: Developmental Trajectories

The use of trajectories in the study of developmental disorders has its origin in growth curve modeling (see, e.g., Chapman, Hesketh, & Kistler, 2002; Rice, 2004; Rice et al., 2005; Singer Harris et al., 1997; Thelen & Smith, 1994; van Geert, 1991) and in the wider consideration of the shape of change in development (Elman et al., 1996; Karmiloff-Smith, 1998). The use of simple correlations to explore relationships between different cognitive abilities in disorders may also be seen as falling within the trajectory approach (discussed later in this article). The impetus to move from matching to trajectory-based studies was a motivation to place development at the heart of explanations of developmental deficits. The phenotype associated with any neurodevelopmental disorder does not emerge full blown at birth but, rather, develops gradually and sometimes in transformative ways with age. For example, when Paterson, Brown, Gsödl, Johnson, and Karmiloff-Smith (1999) gave individuals with WS and Down syndrome a vocabulary task and a number task, they found a different pattern of relative results between the two disorders depending on whether the participants were toddlers or adults. The use of trajectories creates a focus on change over time and discourages static interpretations of developmental deficits as if they represented focal damage to preformed systems (Bishop, 1997; Karmiloff-Smith, 1997; Mareschal et al., 2007; Thomas & Karmiloff-Smith, 2002).

The aim of the developmental trajectory approach is twofold. First, it seeks to construct a function linking performance with age for a specific experimental task. Separate functions are constructed for the typically developing group and for the disorder group, and the functions are then compared. Secondly, it aims to establish the *developmental relations* between different experimental tasks, assessing the extent to which performance on one task predicts performance on another task across development. Once more, the developmental relations found in the disorder group can be compared against those observed in a typically developing group. Trajectories may be constructed in three ways: (a) they may be constructed on the basis of data collected at a single point in time, in a cross-sectional sample of individuals varying in age and/or ability; (b) they may be constructed on the basis of data collected at multiple points in time, tracing longitudinally changes in individuals usually of the same age; or (c) they may combine both methods, with individuals varying in age followed over two more measurement points. In this article, we concentrate on the first type of trajectory, which uses a single point of measurement, although we also show how this method is related to longitudinal and mixed types. Hereafter, where we refer to the developmental trajectories method, unless otherwise specified, we will intend one-time-of-measurement designs for data similar to those collected in the matching method.

Constructing Trajectories

We first consider functions that link performance with CA and the comparisons with typically developing controls that this permits. We then consider developmental relations and functions that link performance with MA, which may serve as a more stringent test of delay/deviance hypotheses.

For a cross-sectional design, the trajectory method works as follows: A disorder group is recruited in which there is a reasonable developmental age range (i.e., spanning childhood, adolescence, and adulthood but not adulthood alone). Performance is assessed on the task that investigates the cognitive process of theoretical interest (which we refer to here as the *experimental task*). Additionally, data are collected on further tasks. These will usually be standardized tests that yield MA-equivalent performance levels for typically developing children.

However, they may also be other tasks that yield behavioral data of potential theoretical relevance to the target domain. Results can be collected on as many measures as are thought relevant to the cognitive process under study (within limits of practicality). A typically developing comparison sample is then recruited that spans from the youngest MA of the disorder group on any of the standardized measures to the oldest CA, and the performance of these comparison individuals is assessed on the experimental task and any additional behavioral measures. Under the assumption that the typically developing group is representative of the typical population, there should be no need for the typically developing group to be assessed on the standardized tests (although these data may be optionally collected to verify that this assumption holds).

The trajectory approach relies on the use of an experimental task that yields sensitivity across the ability range of the disorder group, that avoids floor and ceiling effects where possible (in common with the matching approach), and that has conceptual coherence with the domain under investigation. It is worth noting that the first of these criteria, sensitivity across the ability range of the disorder group, may be one of the hardest to fulfill. This is particularly the case in domains that are characterized by early development, where measures may exhibit ceiling effects at an age when other domains are still showing marked behavioral change over time. An example might be the development of speech compared to vocabulary or syntax. One of the biggest current challenges is to calibrate measurement systems to afford age-level sensitivity while at the same time retaining conceptual continuity over large spans of time (i.e., still constituting measures of the same process at different ages).

Currently, there are few theoretically interesting measures available for studying language disorders that exhibit these characteristics. Instead, sometimes researchers can be tempted to adopt subtests that, despite being psychometrically sound measures, either map only modestly onto interesting linguistic constructs or are overly broad (such as word–picture matching vocabulary tests). One response is to appeal to more sensitive dependent measures such as reaction time. Although reaction times can be noisy, they continue to exhibit developmental change when accuracy levels are at ceiling. A second response is to use implicit rather than explicit measures of performance to assess underlying process. Implicit measures are online, time-sensitive assessments of behavior in which the participants are usually unaware of the experimental variables under manipulation, such as the frequency or imageability of words in a speeded recognition task (Karmiloff-Smith et al., 1998).

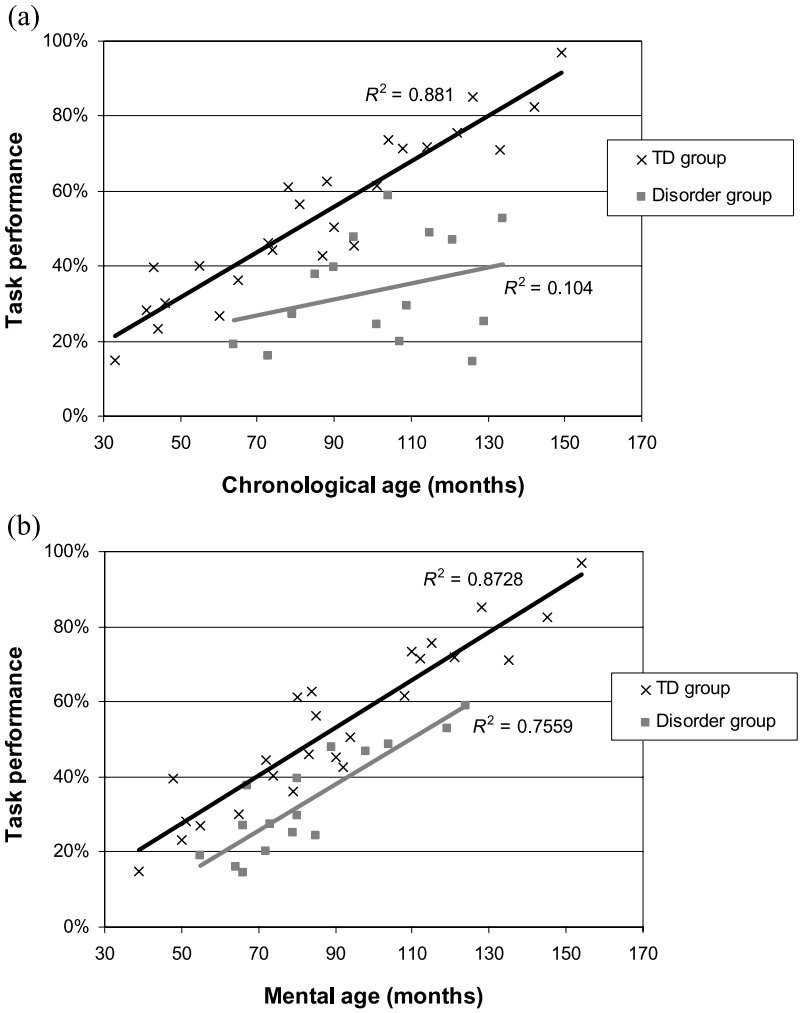
The analysis begins by constructing a task-specific developmental trajectory for the control group, using regression methods to derive a function linking performance on the experimental task with age.[1] We will mostly assume the use of linear methods, since these aid in the understanding of relationships between trajectories (see next section). This may mean transforming either age or the dependent variable or both to improve linearity. Figure 2 shows an illustrative set of results for a typically developing group and a disorder group. The figure depicts all the individual data, reflecting one of our preferences in using the trajectory approach.

There are now three types of comparisons that can be made between the disorder group and the typically developing (TD) trajectory. The first type of comparison is theory neutral. Here, the researcher merely asks whether the performance of each individual in the disorder group on the experimental task fits anywhere on the TD trajectory. If the experimental task has only a single dependent variable, this may not be a particularly useful comparison. That is, if TD performance stretches from 0% to 100% on some measure, it is evident that any individual can be fit on this trajectory. The comparison is, in fact, tantamount to standardizing one's own experimental task, so that an MA measure can be derived for each individual in the disorder group (the mean age of the TD sample at which a given performance level is exhibited). However, when the experimental design includes two or more measures (e.g., performance on high-frequency vs. low-frequency items), the theory-neutral comparison can be much more informative. The researcher can ask whether a given disparity between the two measures (e.g., the frequency effect) for an individual with the disorder can be observed anywhere on the TD trajectory. If it cannot, here is a theory-neutral marker of atypicality. (Strictly speaking, it is theory neutral with respect to the comparison; there is a theory in the experimental design that the relationship between the two measures, such as performance on high- and low-frequency items, should be developmentally robust).

The second type of comparison now allows for the construction of a trajectory for the whole disorder group, linking their performance on the experimental task with their CA. This trajectory can then be compared with the TD trajectory to assess whether the disorder group shows a difference in their developmental performance on the task. It is fairly likely that a difference will be found when studying areas of weakness in the disorder. However, it is a more open question for cognitive domains outside the primary deficit (such as nonverbal abilities in children with SLI). For a single dependent variable, the comparison of two trajectories involves a linear regression model with one between-groups factor. For multiple dependent variables (such as in the example of the frequency effect),

[1]Linear regression may be approximated with ANOVA methods by splitting the age range into several groups and including a multilevel age factor (see Ansari, Donlan, & Karmiloff-Smith, 2007).

Figure 2. Example of data from a developmental trajectory design. TD = typically developing.



this involves a mixed-design linear regression model including within-participants factors to compare several trajectories simultaneously.[2] Confidence intervals around the regression line can be used to assess the age at which trajectories converge or diverge. Figure 2(a) depicts data for the CA-based comparison. Note that the TD group extends to a younger age, and in this case, the disorder group appears to have a lower level of performance and to be developing more slowly. Confidence intervals around the TD trajectory can also be used to assess whether each individual in the disorder group falls outside the range of performance expected for his or her CA.

The third type of comparison considers developmental relations in the disorder group. A separate trajectory is constructed for each standardized test measure collected from the disorder group, in which a function is derived linking the MA (test age equivalent) on that test with task performance. Each MA trajectory can then be compared against the TD trajectory. Importantly, if task performance is in line with a given standardized measure, then plotting the disorder group's data according to each participant's MA should "normalize" the atypical trajectory—that is, move it to lie on top of the TD trajectory.

Developmental relations can also be assessed between the experimental task and any other behavioral measures collected. Each additional behavioral measure is used in turn as a predictor, with the aim of discovering whether the relationship between the behavioral measure and experimental task performance is the same in the TD and disorder groups. The use of simple correlations to explore developmental relations effectively falls within the trajectory approach. However, when using simple correlations, researchers do not always plot these trajectories to illustrate the degree of variability, or

[2]See www.psyc.bbk.ac.uk/research/DNL/stats/Thomas_trajectories.html for sample data and worked examples of trajectory analyses using SPSS.

establish the linearity of relationships between abilities, or check the influence of outliers on the relationship, or check the presence or absence of floor and ceiling effects, and so forth. In our view, the more explicit use of trajectories is, therefore, preferable when relationships are explored.

More sophisticated comparisons are possible. For example, one can use the TD trajectory to standardize the performance of the members of the atypical group. Suppose that the experimental task was some aspect of morphology and one had collected standardized scores for the disorder group on a receptive vocabulary test as a measure of their verbal mental age (VMA). One can then derive a residual score for each individual in the disorder group based on the difference between their observed task score (e.g., on the morphology task) and the score predicted by their MA, according to the TD trajectory (see Jarrold & Brock, 2004). These residuals can be standardized to create *z* scores that can be compared across different experimental tasks. For example, one could also derive *z* scores for the disorder group on a syntax task and ask whether, on the basis of their VMA, there are disparities in the expected levels of morphology and syntax. Here, comparisons are made across different experimental tasks (e.g., morphology, syntax) standardized on the same MA measure (e.g., receptive vocabulary); comparisons are also possible across the same task (e.g., morphology) under standardizations based on different MA measures (e.g., a receptive vocabulary test and a receptive grammar test; see, Jarrold, Baddeley, & Phillips, 2007, for details of these methods).

As long as there is an opportunity to collect multiple standardized test results from the disorder group, the trajectory method gives great flexibility at the analysis stage to evaluate potential relationships to the TD trajectory. This contrasts with the matching approach, where a decision is made at the design stage to recruit an MA-matched control group based on a particular standardized test. The trajectory approach requires only a single TD control group, whereas the matching approach requires as many TD control groups as there are measures of MA. Usually, a larger number of TD controls will be collected in the trajectory approach with a weaker selection bias, giving a fuller picture of typical development on the task. Variability in the disorder group sometimes means that there can be different distributions of CAs and MAs. The trajectory method is tolerant to this difference provided that (a) there is variability in both CAs and MAs in the disorder group and (b) the TD group spans from the youngest MA in the disorder group to the oldest CA—that is, it covers their full ability and age range.

Figure 2(b) depicts performance plotted against an MA measure. For these illustrative data, it becomes evident that the disorder group has a lower level of performance than the TD group even when their lower MA is taken into account. However, the disorder group is now developing at the same rate. A result of this type would suggest that, to the extent that the standardized test is a valid index of development in the target cognitive domain, the delay is uneven across component processes; it is worse for the experimental task than for the standardized task.

The trajectory method is advantageous where there is a wide age (and, potentially, ability) range in the disorder group and experimental task sensitivity exists across this range. These features contrast with the matching approach, which is ideal for narrow age ranges and can tolerate a test with a narrow sensitive range, as long as that range is appropriate for the ability of the disorder and control groups sampled. Many studies that were initially designed using the matching approach supplement the ANOVA-based comparison of the matched groups with regression analyses that explore the relationship of the dependent variable to other dimensions of developmental progress (e.g., MA) and sometimes even CA, as dictated by the trajectories approach. This combination approach offers the advantage of precise matching as well as the consideration of change over time and developmental relations. However, there is a concurrent risk that the sampling of age and ability ranges will not be optimized for either approach.

In common with matching, the trajectory approach seeks to avoid floor and ceiling effects in experimental measures, particularly for the disorder group (see later examples for problems that can arise if floor or ceiling effects are present). Where standardized tests are used to derive MAs, similar caveats apply regarding the way age-equivalent scores mask potential variability in the TD group (McCauley & Swisher, 1984). The similarities and differences between matching and developmental trajectories methodologies are summarized in Table 1.

The developmental trajectories considered thus far are constructed based on a single measurement point and are cross-sectional. Longitudinal studies necessarily give a truer picture of the range of individual development. As we discuss later, when cross-sectional trajectories are constructed, they should be validated by longitudinal follow-up. The trajectory method outlined here should be viewed as one of a suite of statistical methods for studying change over time, along with hierarchical linear modeling and structural equation modeling. Both the latter methods exploit longitudinal data. Hierarchical linear modeling derives linear trajectories for each individual across different measurement points (the "Level 1" analysis) and then compares the intercepts and gradients of the individual trajectories for the TD and disorder groups (the "Level 2" analysis). Where the parameters for the two groups systematically differ (say the gradients are always steeper in the TD group), the method can then be used to assess whether any other experimental measures, such

Table 1. Comparison of the methodologies for investigating developmental disorders.

Methodology	Matching	Developmental Trajectories
Age range	Narrow age range	Wide age range
Comparisons	1. Chronological-age–matched control group 2. Theory-dependent mental-age match (1 control group per theory)	1. Theory-neutral ("can each individual from the disorder group fit *anywhere* on the TD trajectory?") 2. Performance predicted by chronological age 3. Performance predicted by mental age of disorder group (as many comparisons as standardized tests run on disorder group) or by performance on other experimental tasks to derive *developmental relations*
Discrimination	In sensitive range of test (can be narrow) Avoid floor and ceiling effects	In sensitive range of test (must be wide) Avoid floor and ceiling effects
Statistic	Compare group means	Compare group intercepts and group gradients
Aim	Factor out age from comparison	Derive function relating performance to age

Note. The developmental trajectories approach here refers to single-point-of-measure, cross-sectional trajectories.

as those of MA, predict the observed differences (see, e.g., Rice, Wexler, & Hershberger, 1998 and Rice, Tomblin, Hoffman, Richman, & Marquis, 2004, for examples of the method applied to SLI; Chapman et al., 2002, for its application to language development in Down syndrome; Peugh & Enders, 2005, and Willett, Singer, & Martin, 1998, for statistical methods).

Structural equation modeling is a method that can be used to evaluate longitudinal correlation data (e.g., performance on Task A at Time 1 vs. performance on Task B at Time 2) against hypotheses concerning the underlying causal relationships between abilities. For example, this method has been used in the study of reading deficits to evaluate the relationship between phonological awareness, vocabulary knowledge, and letter knowledge during reading acquisition in typical children and children with dyslexia (see, e.g., Torppa et al., 2007; see Curran & Hussong, 2002, for statistical methods). Structural equation modeling focuses on theory testing rather than theory development, as it evaluates the fit of a proposed causal model against the data rather than discovering these models from the data themselves. It can be used in combination with exploratory factor analysis, which can generate potential structures to test.

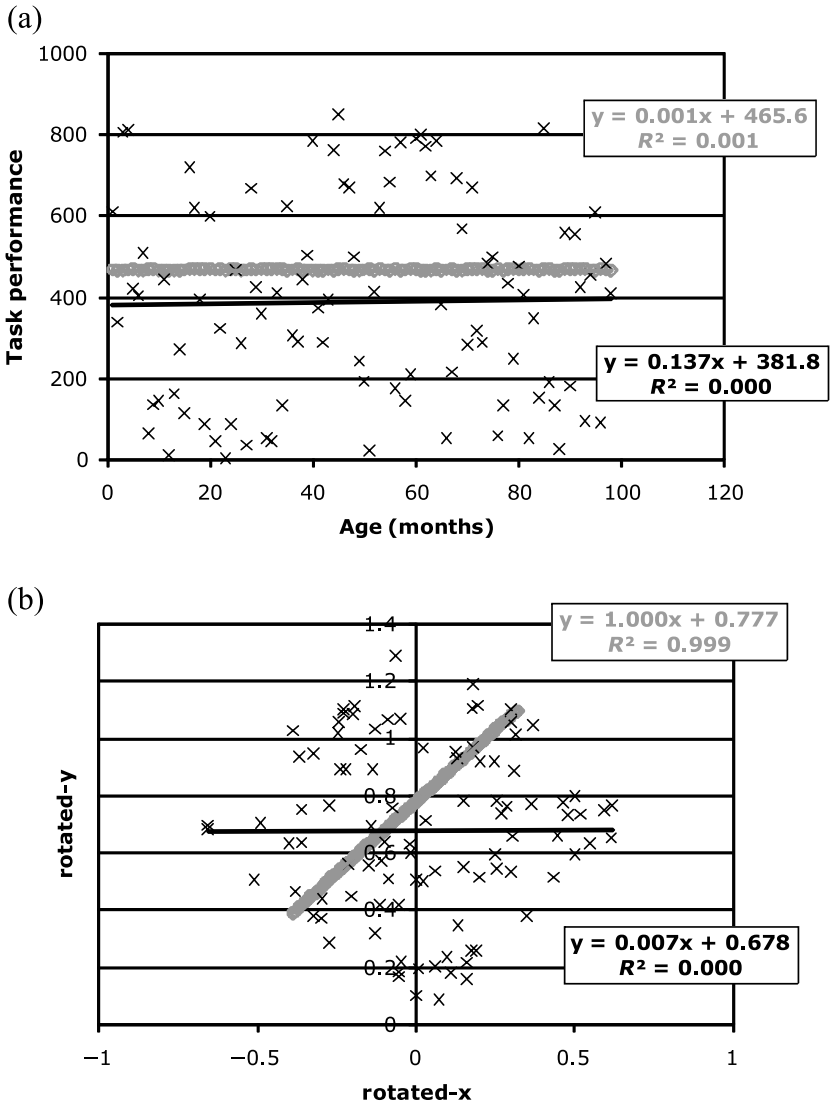
Using Trajectories to Distinguish Types of Developmental Delay

We are now in a position to consider how trajectories may be useful for studying developmental delay. Under the matching approach, a cognitive ability in a disorder group is described as delayed if performance falls below the CA-matched control group but resembles that of a control group matched on a MA measure deemed relevant for the target cognitive domain. The thrust of this section is that, when construed in terms of developmental trajectories, the performance of the disorder group can resemble that of the younger TD group in more than one way. We believe that one of the reasons neurocognitive explanations of developmental delay are thin on the ground is that "delay" is not sufficiently detailed as a descriptive term. In this section, we show how the use of trajectories distinguishes at least three forms of delay, and we show how additional descriptors also discriminate patterns of development that may index different underlying causal mechanisms.

The first step in trajectory analysis is to establish that there is a reliable relationship between performance on the experimental task and CA in the typically developing group. In other words, it is necessary to demonstrate that we have chosen an experimental measure that is sensitive to developmental change. Assuming that linear regression analyses have established this relationship, our next step is to evaluate how performance changes with age in the disorder group. But what if we find that there is no reliable relationship between task performance and age in the disorder group?

One difficulty with the linear regression analysis is that a nonsignificant relationship may arise under two conditions: (a) when the distribution of performance scores is random with respect to the predictor of age and (b) when the points are distributed horizontally (see Birdsong, 2005, for a similar point). In both cases, age is not useful in predicting performance. However, in the second case, provided that the task measure is in the sensitive range and it has been established that the TD group improves across the age range, it may be that individuals with the disorder have indeed progressed as far as they can given the constraints of their cognitive systems and that the trajectory has a gradient of zero. That is to say, the absence of a reliable relationship in the disorder group

Figure 3. Simulated data of (a) two nonreliable trajectories with different variance around the regression line and (b) the two trajectories after 45° anticlockwise geometric rotation—only the trajectory with small variance becomes reliable, implying that it genuinely had a zero gradient.



may be a real reflection of the fact that their performance is static across age. Figure 3(a) illustrates idealized versions of the two cases, which we refer to as *no systematic relationship* and *zero trajectory,* respectively. In both cases, the best-fit regression lines are flat; for one trajectory, the best-fit line lies in the middle of a random data cloud; for the other, the points are tightly clustered around a narrow performance range across development.

Where null trajectories appear in the disorder group but not the TD group, we have found it useful to distinguish between the two null cases by using a rotation method. Figure 3(b) depicts the same data but transformed by a 45° anticlockwise rotation in geometric space. When the analyses are repeated on the rotated data, the zero trajectory now produces a highly significant regression (the R^2 value changes from .0011 to .9999 following rotation), whereas the no systematic relationship produces a similar degree of fit before and after rotation (R^2 changes from .00030 to .00004). A trajectory that switches from a nonsignificant R^2 to a significant R^2 following rotation is suggestive of a zero trajectory rather than no systematic relationship. This method relies on the availability of a typical developmental trajectory to provide a benchmark of expected performance variability around the trajectory.[3] When a zero trajectory is observed, the

[3]See www.psyc.bbk.ac.uk/research/DNL/stats/Thomas_trajectories.html for further details of this method and a worked example.

explanation for this pattern would remain to be articulated. Floor effects ceiling effects would need to be ruled out before one inferred that flat performance reflected the limitations of internal cognitive constraints on development.

Assuming that we have two reliable linear trajectories, one for the TD group and one for the disorder group, these trajectories can now be statistically compared. The test indicates whether there is a significant difference in the rate (gradient) and/or the onset (intercept) of the trajectories. Importantly, where there is a difference between the two trajectories, three different types of descriptive delay can now identified. These are depicted in Figure 4 with illustrative data. In Figure 4(a), there is a significant difference in the intercept. Here, delay is manifested in a later onset of development. In Figure 4(b), there is a difference in the gradient between the two trajectories. Here, delay takes the form of a slowed rate of development in the disorder group. In Figure 4(c), there is a difference in both parameters, implying that development has both a delayed onset and a slowed rate.

A focus on trajectories allows further descriptors to be attached beyond delay, which may ultimately index different underlying developmental pathways. For example, as we suggested above, the TD group may exhibit a reliable trajectory, but the disorder group may exhibit no reliable change in performance with age. Figure 4(d) and Figure 4(e) illustrate two further types of difference. In the first, a linear relationship is observed in the TD trajectory, but a nonlinear trajectory is observed in the disorder group. In the second, a linear relationship is observed in the TD trajectory, and this is initially tracked in the disorder group, but the disorder group then asymptotes at a lower level of performance.

These alternative descriptors are assigned when an alternative function gives a significantly closer fit to (i.e., a better explanation of) the data than the linear equation. The R^2 value for a regression model indexes how well the model fits the data (specifically, the proportion of variance explained), and R^2 values can be derived for different functions fitted to the same data (e.g., in the SPSS Regression Curve Estimation facility). A higher R^2 gives a better fit. To illustrate, linear and nonlinear functions were fit to the disorder trajectory in Figure 4(d). The linear function produced an R^2 of .900, whereas the logistic function (an s-shaped curve) produced an R^2 of .990. Therefore, one might view the disorder trajectory as nonlinear. It is possible to test whether one function is a statistically significantly better fit than another function by discounting for the extra parameters available in the more complex equations. For example, when linear and nonlinear functions were fit to the data in Figure 4(e), the linear function produced an R^2 of .943, whereas a quadratic function (including a variable of age-squared) produced an R^2 of .998. Using the extra sum-of-squares test for comparing nested models (see Motulsky & Christopoulos, 2004), a statistical comparison indicated that the quadratic was a reliably better model, $F(1, 2) = 70.1$, $p = .014$. The disorder trajectory would, therefore, be classified as nonlinear and, given its shape, as exhibiting a premature asymptote. Finally, because nonlinear functions also have intercepts, one can characterize a trajectory as separately showing a delayed onset followed by a nonlinear trajectory.

How would the matching approach deal with the different types of delay we have described? The illustrative data in Figure 4 allow us to make this comparison by averaging across groups. Figure 4(f) demonstrates the mean performance of the TD group and the disorder groups with each type of trajectory, collapsed over age, as would be the case in a group comparison. Delayed onset + slowed rate (see *x*-axis in Figure 4[f]) produces the lowest mean score and premature asymptote produces the highest, whereas delayed onset, slowed rate, and nonlinear all produce similar scores. The fact that, from the perspective of the matching approach, some of these groups are indistinguishable suggests that, for wide age ranges at least, the use of trajectories provides a descriptively more powerful empirical vocabulary.

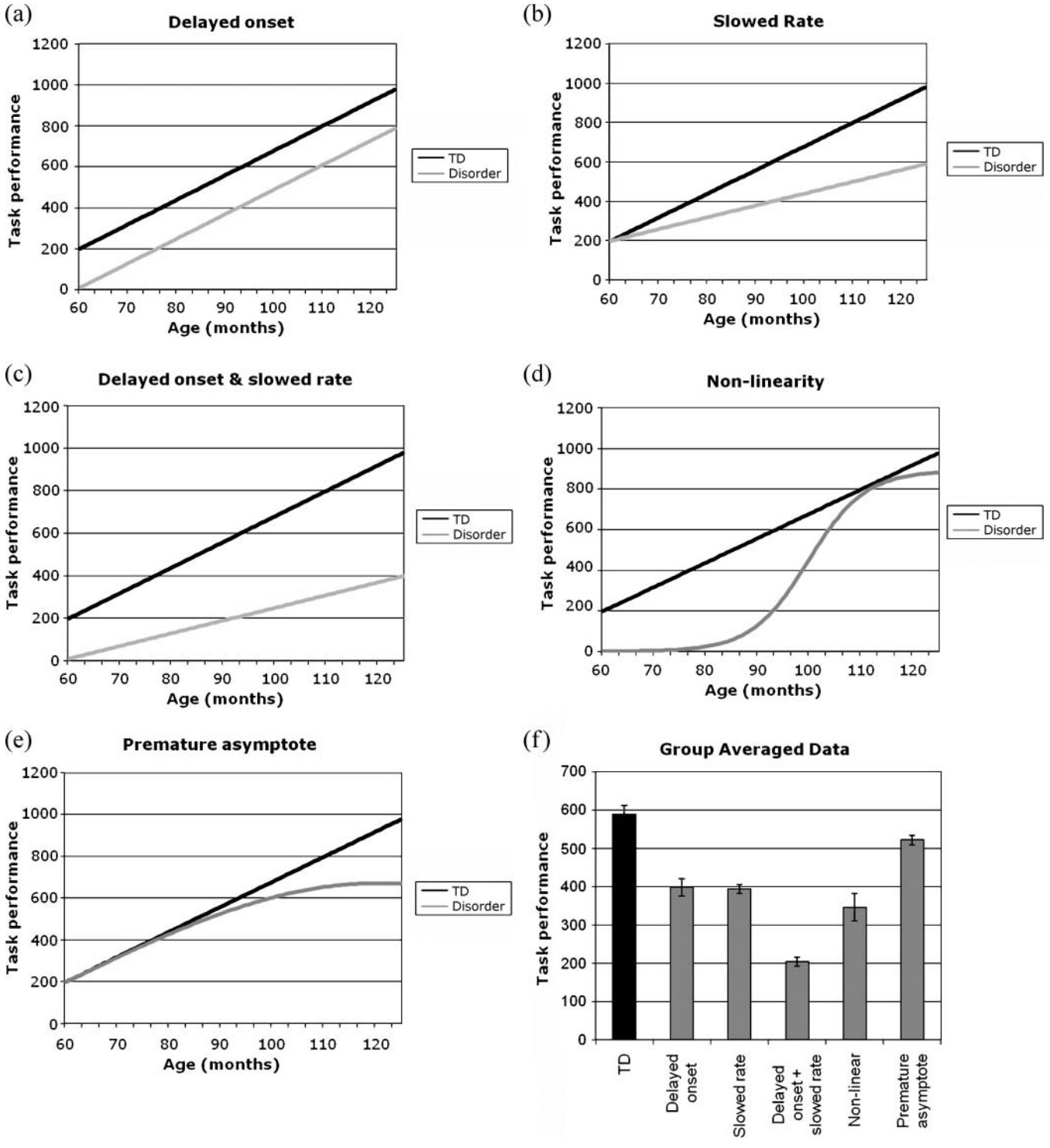
Let us amplify this point. Where the matching approach can encourage a monolithic descriptive partition between "delay" and "deviance," the use of trajectories distinguishes at least seven ways that a disorder group can statistically differ from a control group in the functions that link performance and age (or MA): (a) delayed onset, (b) slowed rate, (c) delayed onset + slowed rate, (d) nonlinear, (e) premature asymptote, (f) zero trajectory, and (g) no systematic relationship with age. An accurate characterization of patterns of change is, of course, a necessary precursor to formulating causal accounts of developmental impairments.

This richer taxonomy of developmental delay, with its focus on developmental change and developmental relations, draws similar conclusions to the recent work of Rice and colleagues (see, e.g., Rice, 2004; Rice et al., 2005). For comparison, Rice (2004) suggests that developmental trajectories should be characterized in terms of their onset timing, their acceleration rate, and points of change in their acceleration and that separate trajectories should be established for the delineated subcomponents of the linguistic system. Rice et al. (2005, p. 22) place particular emphasis on the utility of onset differences in language development, arguing that delayed onset may be a hallmark characteristic found across most of the known clinical forms of language impairments.

Examples of the Trajectory Approach

In this section, we describe four examples of studies that have used the single-point-of-measure trajectory

Figure 4. The shape of delayed (a–c) and atypical (d–e) developmental trajectories, along with the same data plotted in terms of group means (f) on an experimental task (*y*-axis scale is arbitrary).



method to explore potential differences between one or more developmental disorder groups and a TD control group. These examples focus either on language development or on the developmental relations between verbal and nonverbal development. They serve to illustrate a number of methodological points that arise in using the trajectory approach.

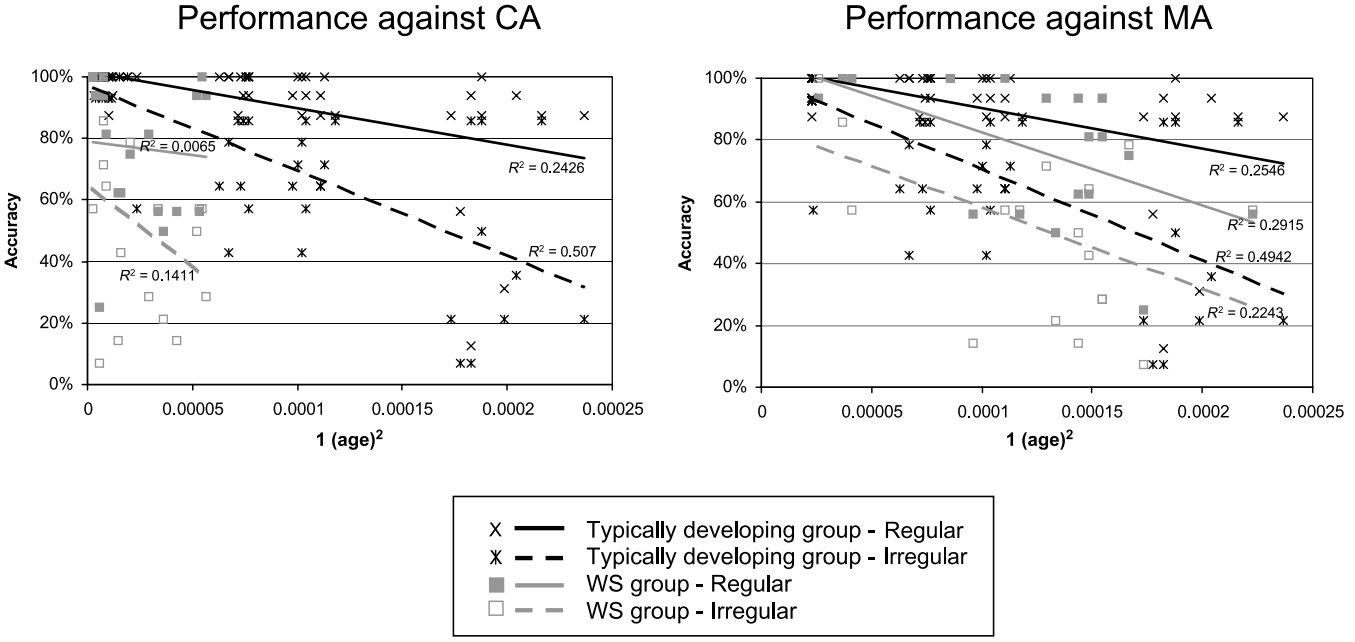
Example 1: Inflectional Morphology in Williams Syndrome

Early published and unpublished studies of language development in WS suggested that the individuals with this disorder might have greater problems inflecting irregular nouns and verbs than regular nouns and verbs (Bromberg et al., 1994; Clahsen & Almazan, 1998). This is of theoretical interest because performance on inflecting regular and irregular items is taken to index either the involvement of different mechanisms (rule-based vs. associative learning mechanisms) or the influence of different information sources (phonological vs. lexical-semantic), depending on the theory (see Thomas & Karmiloff-Smith, 2003, for a review). However, these initial studies were compromised by small participant numbers and/or the absence of appropriate statistics. Moreover, the most salient characteristic of language development in WS is that its onset is delayed (see, e.g., Meyer-Lindenberg, Mervis, & Berman, 2006). One characteristic of typical development is that irregular inflections are harder to learn than regular inflections. Might, then, the apparent problem in irregular inflection stem from a delayed onset in language development rather than a specific deficit to some component of the language system?

This question is amenable to study by a matching approach that compares the mean performance of a WS group in producing regularly and irregularly inflected forms with the mean performance either of a CA-matched or an MA-matched control group, respectively. Figure 5 shows the results of a study that used the trajectory approach. This study sought to capture the change in accuracy levels in a past tense elicitation task with increasing CA or increasing VMA across a wide range of both measures. The data are from 18 individuals with WS and 46 TD controls (Thomas et al., 2001). Two groups and two verb types produced four trajectories, which were analyzed with a mixed-design linear regression model. In this case, the data indicated that when the trajectories were plotted by CA, there was a greater deficit for irregular verbs than regular verbs in the WS group compared with the TD group (i.e., a Group × Verb Type interaction). However, there was no such deficit when the trajectories were plotted by MA. In other words, irregular verb performance was in line with the development of the language system, as indexed by the standardized test used.

This example illustrates several methodological points. First, the use of scatter diagrams and best-fit lines in the trajectory approach makes explicit the degree of variability present in both disorder and TD groups as

Figure 5. Past tense elicitation performance for typically developing (TD) and Williams Syndrome (WS) groups, for regular (talk) and irregular (drink) verbs, plotted against chronological age (CA). Mental age (MA) was measured using a test of receptive vocabulary (the British Picture Vocabulary Scale, Second Edition [BPVS-II]; Dunn, Dunn, Whetton, & Burley, 1997). (Data from Thomas et al., 2001.)



well as the proportion accounted for by the trajectory. Second, the data were partially compromised by a ceiling effect in more able participants (especially in the control group), a problem that has affected many studies of inflectional morphology in developmental disorders (Brock, 2007). In an attempt to address this problem, the data were linearized by plotting performance against $1/(\text{age})^2$, where age was calculated in months, but clearly it would have been preferable if the test had been in the sensitive range for all participants. Third, comparison of the two panels of Figure 5 makes clear that the distribution of CAs and MAs was different in the disorder group. Analytically, this is not problematic provided that there is variability in both dimensions and that the TD group extends from youngest MA of the disorder group to oldest CA. Last, standardized tests usually have a maximum age (in this case, 17 years and 6 months). This presents a difficulty in comparing the disorder group against TD at older CAs because, obviously, no individual can produce a test age above the ceiling for the test. If the disorder group never reaches ceiling on the standardized test, the difficulty is, to some extent, resolved by assigning an MA of the ceiling value to any individual in the TD group whose age falls above the ceiling. For comparable work on the development of regular and irregular past tense morphology in cases of specific and nonspecific language impairment using longitudinal data and growth curve modeling, see Rice et al. (2004).

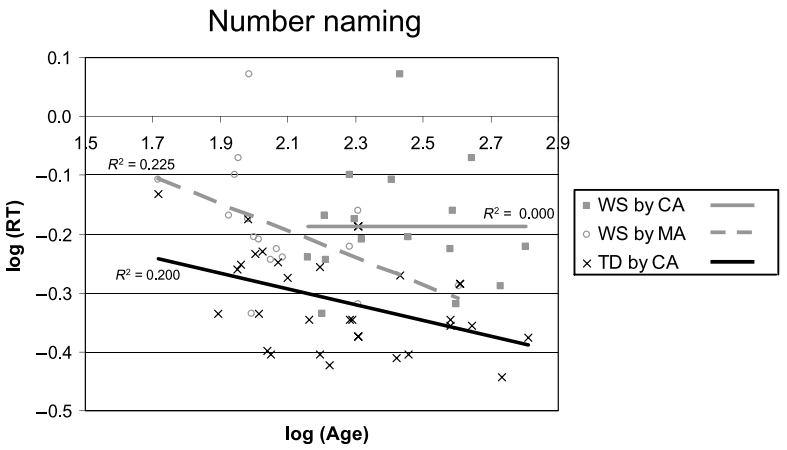
Example 2: Picture Naming in Williams Syndrome

Early work on language development in WS also made another interesting claim. Following anecdotal reports of the presence of rare or unusual words in the spontaneous language of individuals with WS, some researchers suggested that this behavior reflected atypical structure in their lexicon and in particular, an attenuated encoding of word frequency (Rossen, Klima, Bellugi, Bihrle, & Jones, 1996). Thomas et al. (2006) explored picture-naming reaction times in a sample of 16 individuals with WS and, once more, compared these times to those of a TD trajectory ($n = 16$). Pictures varied according to semantic category (object, action) as well as frequency.

When developmental relations were explored using a mixed-design linear regression model, this study indicated that the frequency effect on picture naming in the WS group was in line with their VMA. No atypicality was present. Several points are of interest here. These are perhaps more clearly illustrated by a measure of baseline naming speed in the two groups derived by measuring naming times for the numerals 1–9. These highly familiar, overlearned stimuli were named with 100% accuracy in both disorder and TD groups. Naming times are shown in Figure 6.

First, for tasks in which accuracy is at ceiling, more sensitive measures such as response time can still reveal group differences. Moreover, properties of the underlying systems may be explored by the manipulation of implicit variables, such as frequency and semantic category in the current example. The variables are implicit in the sense that one would normally expect the participants to be unaware of them given the task definition (here, naming) and, therefore, unlikely to deploy differential, explicit strategies for the different values of each variable (e.g., for high- vs. low-frequency words). Notably in this case,

Figure 6. Naming times for numerals 1–9 for TD and WS groups, plotted against CA (log-log transformed). MA was measured using the BPVS-II (Dunn et al., 1997). This figure is reproduced with permission from Taylor & Francis Ltd, www.informaworld.com, publisher of the journal *Language and Cognitive Processes* (see Thomas et al., 2006, for full reference).



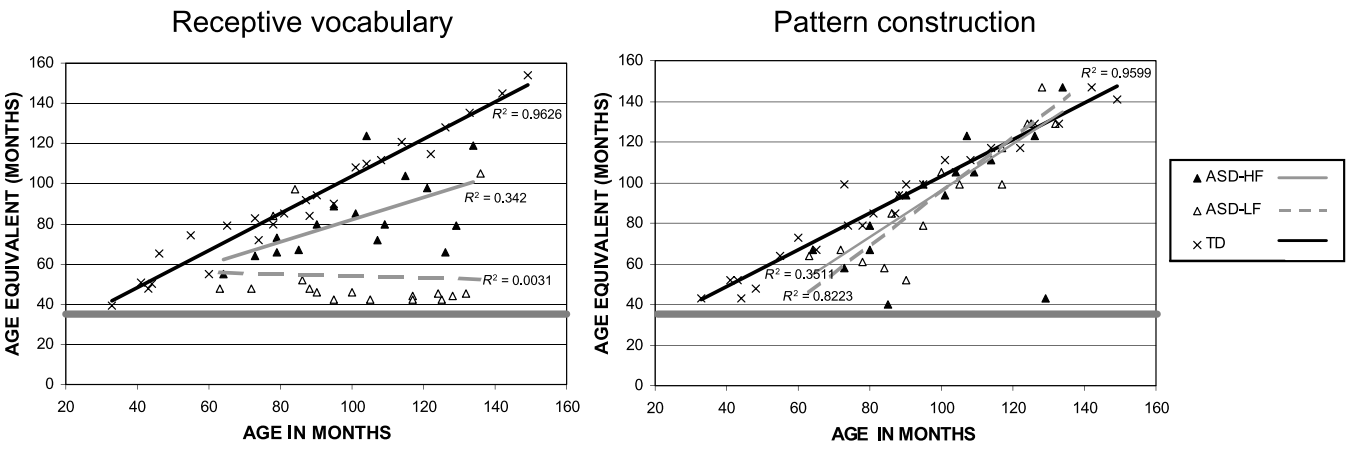
although the WS group proved slower than expected both for their CA and their MA, the implicit variables produced the same modulation of response times in both groups, suggesting similar underlying structures. Second, in the TD population, reaction times tended to decrease with expertise according to a power law (Cohen, Dunbar, & McClelland, 1990). In this case, therefore, a log-log transform was required to linearize the data for the disorder and TD groups. Third, in the cross-section sample, there was no reliable relationship between naming speed and CA in the WS group. However, there was a significant relationship between naming speed and VMA. This is a common result for cross-sectional studies of disorder groups. The poor predictive power of CA must be interpreted with caution, an issue to which we return shortly.

Example 3: Verbal Versus Nonverbal Abilities in Autism: Spectrum Effects

Cross-syndrome cross-domain comparisons can be very informative about the atypical constraints operating in developmental disorders. One can begin by making simple comparisons based on the multiple subtests of standardized intelligence tests (although one must acknowledge that in some respects, these tests have limited sensitivity; see Karmiloff-Smith, 1998; Karmiloff-Smith et al., 1998). Figure 7 depicts data taken from Annaz (2006) in her comparison of WS, Down syndrome, and autism for children between 5 and 12 years of age. Notably, Annaz (2006) collected data from low-functioning as well as high-functioning children with autism in order to explore the influence of the spectrum of this disorder (see Annaz, Karmiloff-Smith, Johnson, & Thomas, 2009). High-functioning (n = 16) and low-functioning (n = 17) children were assigned to their groups according to the Childhood Autistic Rating Scale (CARS; Schopler, Reichler, & Rochen, 1993). Figure 7 plots test ages derived from a verbal test of receptive vocabulary (the British Picture Vocabulary Scale, Second Edition [BPVS-II]; Dunn, Dunn, Whetton, & Burley, 1997) and from a nonverbal test of visuospatial construction (the Pattern Construction subtest of the British Abilities Scales, Second Edition [BAS-II]; Elliott, Smith, & McCulloch, 1996) against CA for TD (n = 25), high-functioning groups with autism (ASD-HF), and low-functioning groups with autism (ASD-LF). The gray horizon line represents floor performance on each test.

As expected, the TD group received test ages very close to their CAs (R^2 = .9626 and .9599 for the two tests, respectively). For receptive vocabulary, the ASD-HF group produced a reliable trajectory that was slightly lower (i.e., later in onset) than the TD trajectory, although this difference did not reach significance. By contrast, for the ASD-LF group, no reliable trajectory emerged, and indeed most of these children were at or close to floor on the vocabulary test. In one sense, this is not surprising, since one of the markers of severity in autism is the level of language development. However, one might even question whether these data are valid: Perhaps the ASD-LF group was simply unable to complete this task given their low ability level? Figure 7 (right panel) allows us to address this question. These data reveal the developmental trajectories on the BAS-II Pattern Construction task, in which the children are asked to complete geometric puzzles. Here, both groups with autism produced trajectories overlapping with those of the TD group, and indeed the ASD-LF group produced a tighter trajectory than the ASD-HF group (ASD-LF, R^2 = .8223; ASD-HF, R^2 = .3511).

Figure 7. Comparison of test age scores for the BPVS-II (Dunn et al., 1997) and the pattern construction subtest from the British Ability Scales, Second Edition (BAS-II; Elliott et al., 1996), plotted against CA, for typically developing children (TD), high-functioning children with autism (ASD-HF), and low-functioning children with autism (ASD-LF). (Data from Annaz, 2006.)

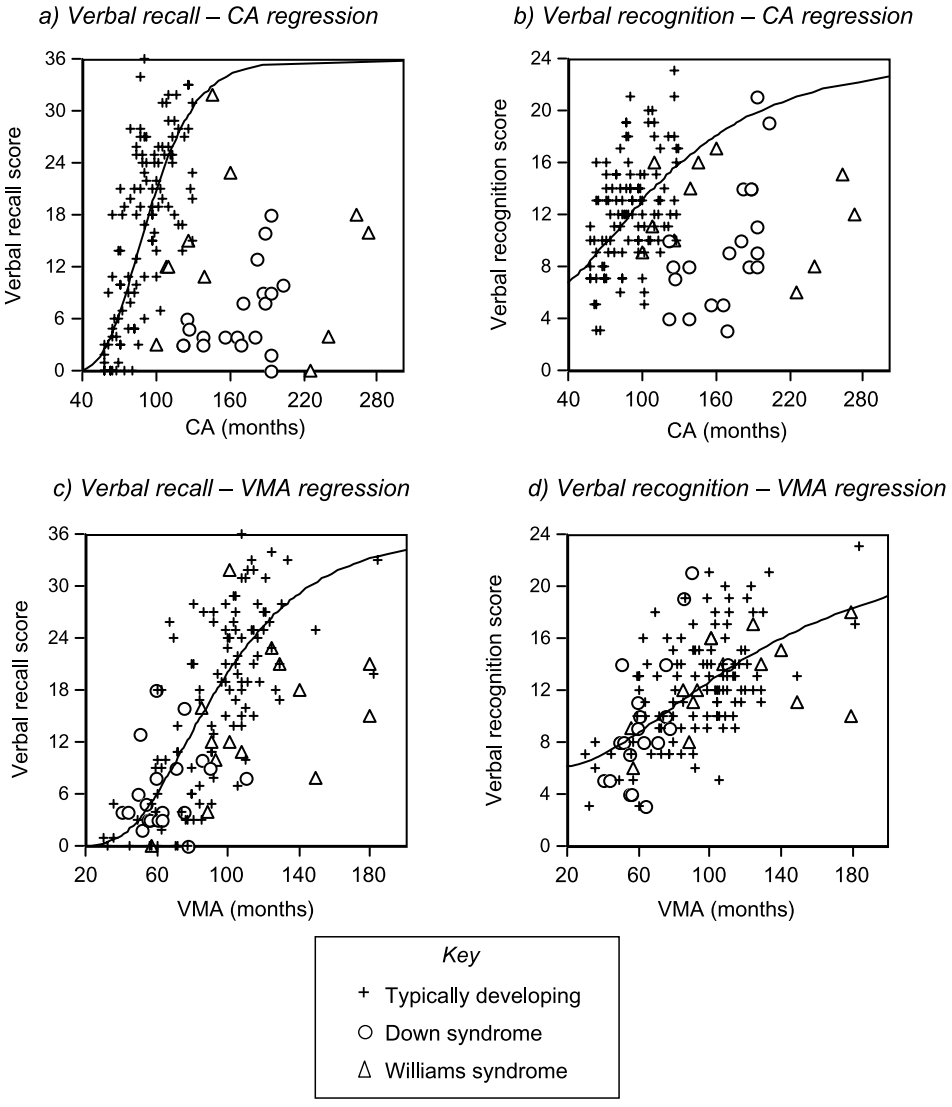


These data revealed stark differences in the profile of children at different points of the autistic spectrum. Methodologically, this example demonstrates that cross-domain comparisons can shed light on the validity of the respective trajectories. The normal profile on pattern construction for the ASD-LF group increases confidence that the lack of improvement on vocabulary in this group is a real phenomenon. Lastly, the two ASD-LF trajectories on the verbal and nonverbal measures illustrate what a developmental dissociation looks like within the trajectory approach. By comparison, in the matching approach, the dissociation would be represented by two scores corresponding to the mean performance of the group on each test (see Karmiloff-Smith, Scerif, & Ansari, 2003, for a discussion).

Example 4: Verbal and Visuospatial Memory in Williams Syndrome and Down Syndrome

In the final example, we consider some more sophisticated techniques to compare developmental relations between abilities in two disorders, again using cross-syndrome cross-domain comparisons. Jarrold et al. (2007) compared the performance of individuals with Down syndrome (n = 20) and individuals with WS (n = 15) to that shown by 110 TD children on the Doors and People Memory Test (Baddeley, Emslie, & Nimmo-Smith, 1994), a measure of verbal and visuospatial recall and recognition memory. Figure 8 plots the performance of these groups on two of the tasks in the battery: the verbal recall

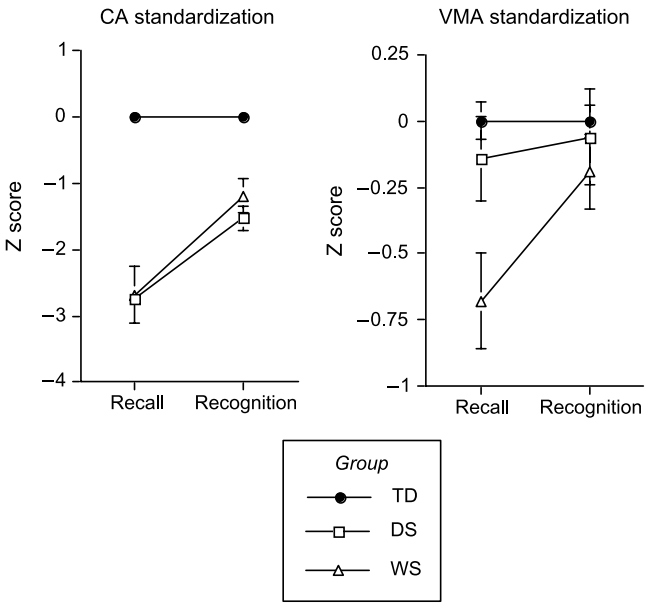
Figure 8. Nonlinear developmental trajectories for verbal recall and recognition tests (data from Jarrold et al., 2007). Verbal mental age (VMA) was measured using the BPVS-II (Dunn et al., 1997).



and verbal recognition tests. The top two panels of the figure show performance plotted against CA, whereas the lower two panels show performance plotted against VMA (again, assessed via the BPVS-II; Dunn et al., 1997). Because of the range of ages and abilities within the TD group, both floor and ceiling effects are apparent, and so the development of performance with age or ability is not linear in this group. Consequently, these regressions were linearized by converting each individual's score into a probit score (the *z* score corresponding to that individual's score on the task as a proportion of the maximum possible) and then regressing that value against the log of either CA or VMA. This produced reliable linear fits, and these in turn allowed the authors to determine the extent to which each individual in the disorder groups exhibited performance that was in line with their CA or VMA (specifically, the residual scores for each individual were standardized on the basis of these linearized regressions).

Figure 9 shows the resultant standardized residual values under the two different forms of standardization and indicates how far each disorder group fell below the normal range for recall and recognition. Three key points can be drawn from these data. First, they further emphasize the fact that atypical groups tend to perform poorly on CA standardizations because their abilities lag behind age-expected levels. A comparison of the scales of the two graphs in the figure shows that when performance is standardized for VMA, the disorder groups are much less impaired. Second, when the two disorder groups are standardized for age, they perform similarly, yet when compared with TD individuals on the basis of VMA, the individuals with Down syndrome are clearly less impaired than those with WS. This reflects the fact that VMA is a relative strength in WS and something of a weaker area in Down syndrome; consequently, broadly comparable overall levels of task performance represent different levels of impairment relative to VMA in the two groups. Finally, the figure shows that the type of regression employed to standardize the data has implications for the interpretation of the results. When the groups are standardized relative to CA, both groups perform poorly on both the recall and recognition tasks. However, under the VMA standardization, the individuals with Williams syndrome show impaired performance on the verbal recall task only. This difference in patterns of impairment reflects the fact that the two tasks are related to CA and VMA in different ways than in the typical standardization sample (see Figure 8).

Figure 9. Verbal recall and recognition performance standardized for CA or VMA according to the BPVS-II (Dunn et al., 1997). (Data from Jarrold et al., 2007.) DS = Down syndrome.



Other studies using the trajectory methodology can be found in Annaz et al. (2009); Brock and Jarrold (2004, 2005); Brock, Jarrold, Farran, Laws, and Riby (2007); Cornish, Scerif, and Karmiloff-Smith (2007); Jarrold, Cowan, Hewes, and Riby (2004); Jarrold et al. (2007); Karmiloff-Smith et al. (2004); and Scerif et al. (2005).

Practical Issues of Using Trajectories

In this section, we briefly expand on three practical issues related to trajectory analysis: interpreting null results, validating cross-sectional trajectories with longitudinal follow-up, and identifying atypicality.

Interpreting Null Results

In some of the examples described previously, there were conditions where no reliable trajectory was found in the disorder group—that is, the function linking age and performance did not pick up a statistically significant amount of the variance. What does it mean when there is no systematic relationship? Does it really mean that performance does not improve with age in the disorder? This would be a pattern that radically departs from the expectations of normal development. However, although it could be true given the data, one has to be cautious with this interpretation. One explanation of the null result for trajectories constructed against CA is that this is simply an artifact of the cross-sectional design. Most disorders show a good deal of variability in how severely each individual is affected. When constructing a cross-sectional sample, there will not necessarily be a relationship between how severely each individual is

impaired and how old they are (and, indeed, one hopes there will not be—to have, say, all the younger children more severely impaired than the older children would represent a recruitment bias). However, a decorrelation between severity and age means that any relationship between age and performance may be weakened or eliminated in the sample; by contrast, severity is factored into the MA, so this is likely to be more predictive of behavior. Given a null CA-based trajectory, it is nevertheless possible that were each individual followed longitudinally, he or she would show improvement.

If a null result is found in a relationship between performance and MA, there are several follow-up questions that must be asked. First, were individuals in the disorder group able to understand the demands and carry out the test, given their level of ability? If one identifies a comparable task in which the same group shows a reliable trajectory (as in the example of low-functioning children with autism), this increases confidence that the nonreliable trajectory for the first test is real. Second, the influence of floor or ceiling effects may also destroy a relationship between performance and age. If the disorder group scores in the sensitive range of the test, this also increases confidence that the nonreliable trajectory is real. Third, assuming that the TD trajectory is satisfactorily linear, it may be that a nonlinear trajectory is appropriate for the disorder and may predict a significant amount of the variability.

Some trajectories can be reliable (statistically significant) but predict a very small amount of the variance, so that performance increases only slightly across the age range sampled. Here, the trajectory approach is beneficial because it necessarily emphasizes the difference between effect size (the size of the intercepts and gradients) and statistical significance. This distinction is sometimes de-emphasized in matching designs that aim to identify delay or deviance based on finding significant differences between the disorder group and CA or MA control groups. Of course, one may legitimately ask what is a sufficient amount of variability for a trajectory to pick up before it should be taken seriously. For example, Figures 2–9 depict significant trajectories whose R^2 values vary from .09 to .98. The answer to this question is that it depends on the effect size that one is expecting given the theory, given the experimental paradigm, and given the existing literature. The poorest fitting trajectories in our examples arose when performance was predicted by CA rather than MA; when performance was close to floor or ceiling, reflecting limits on test sensitivity; and when reaction time data were used that are intrinsically more noisy.[4]

[4]Although reaction time data are noisy, they are nevertheless a developmentally sensitive measure in that they continue to show developmental change when accuracy has reached ceiling.

Given the likelihood that in many cases, the trajectory linking performance and CA for the disorder group will fall below that for the TD group, and given the problems of variations in severity destroying the relationship in cross-sectional analyses, one might ask why it is worth building CA-based trajectories for disorders. Why not jump straight to considering developmental relations in our studies and simply construct trajectories against MA?

There are four reasons why we believe that CA-based trajectories are an important preliminary step in characterizing a disorder. First, there will be abilities on which we do not necessarily expect individuals with disorders to score more poorly (e.g., nonverbal skills in children with dyslexia or SLI). In these cases, the CA trajectories should coincide with the TD trajectory and be statistically different from the CA trajectories in areas of weakness. Second, CA trajectories are theory-neutral descriptions of how performance tends to improve, on average, with age in a disorder (subject to the limitations of cross-sectional designs). By contrast, MA-based trajectories are theory-dependent. Third, by definition, the study of developmental relations focuses on relative abilities, and this may mask absolute differences in comparison to typical development. For example, it has been argued that in WS, the developmental relation between mean length of utterance (MLU) and syntactic complexity is normal (i.e., not significantly different from the TD trajectory for this relation) and, therefore, that language development is itself normal in the disorder (in contrast to, say, Down syndrome, where syntactic complexity is lower than expected given MLU; Mervis et al., 2000). However, it is all too easy to focus on the normality of the relations and ignore the absolute patterns that indicate that the most salient feature of language development in WS is delayed onset (i.e., the WS CA-based trajectory is significantly different to the TD trajectory in its intercept), with some additional suggestion of a premature asymptote (Grant, Valian, & Karmiloff-Smith, 2002; Zukowski, 2001). Last, the comparison of CA- and MA-based trajectories is important to avoid being seduced by novel developmental relations in disorders. For example, let us say that two abilities, A and B, are correlated in a cross-sectional disorder sample but not in the TD sample (e.g., language and verbal memory ability in children with WS; see Meyer-Lindenberg et al., 2006). This could be because Abilities A and B are causally related in the disorder but not in TD. However, it could also occur because Abilities A and B are both constrained by disorder severity (a common causal factor) in the disorder, a factor that does not operate in the TD sample. For these reasons, then, we believe that the study of developmental relations in disorders must be complemented by the initial construction of task-specific CA trajectories.

Validating Cross-Sectional Designs With Longitudinal Follow-Up

We have been clear throughout this article that longitudinal designs are superior to cross-sectional designs for studying development. Our thrust has been that cross-sectional trajectories have some advantages over matching methods in their focus on change over time and the flexibility that they permit in the comparisons that can be made between disorder and TD groups. However, longitudinal designs also have disadvantages. They are costly, place a burden on participants, suffer relatively high dropout rates, and produce long lags between the start of a project and the report of final results.

A more time-efficient and cost-efficient design begins by constructing a cross-sectional study and then uses longitudinal follow-up of some or all of the participants to validate the trajectories predicted by the initial study. This design permits immediate reporting of provisional results, followed by validation of those results in a longitudinal design that is more tolerant of participant dropout. Such longitudinal follow-up can also reveal limitations in the cross-sectional trajectories arising from shortcomings in test sensitivity, such as floor effects. For example, Figure 10 depicts two cross-sectional trajectories for a sample of 28 children with WS between the ages of 5;5 (years;months) and 12;1, plotting test age on a verbal measure (the BPVS-II; Dunn et al., 1997) and test age on a nonverbal measure (the Pattern Construction subtest of the BAS-II; Elliott et al., 1996) against CA. These trajectories replicate a pattern often observed with WS, showing a marked disparity between the development of receptive vocabulary and visuospatial construction skills. Descriptively, the results indicate that receptive vocabulary has a delayed onset and is developing at only a marginally slower rate, whereas pattern construction has both a delayed onset and a severely slowed rate. Some years after these data were collected, we revisited a small subset of 4 of these children, after a delay of between 27 and 49 months. The repeated measures are indicated in Figure 10 with unfilled symbols; thin lines link each follow-up measure to the first measure.

We can now evaluate whether the longitudinal trajectories of these 4 children fall within the confidence intervals predicted by the initial cross-sectional trajectory. The results on vocabulary development are in the affirmative. The only individual who falls below the predicted trajectory on follow-up also fell below it to begin with—this child had a more delayed onset than average but had the same rate. By contrast, the pattern construction findings imply that the initial trajectory was incorrect. Two of the children who were at floor to begin with remained at floor, but the other 2 children showed increases in performance at a much faster rate than predicted; indeed, the rate was comparable to vocabulary development. The follow-up data suggest that the initial pattern construction trajectory mistakenly averaged together floor effects with real developmental improvement. Were the results of the latter 2 children with WS to be representative, the implication would be that the true delay is one that has an impact mainly on onset within the age range studied and that the children with WS vary in the severity of their delays in onset. A more detailed consideration of the use of longitudinal trajectories to validate earlier cross-sectional findings for vocabulary development and pattern construction in WS can be found in Jarrold, Baddeley, Hewes, and Phillips (2001). In the current context, the more general lesson is that trajectories should only be built using scores that are above floor and below ceiling.

Identifying Atypical Development

In matching designs, if the mean performance of the disorder group is the same as the MA-matched control group, the disorder group is classed as exhibiting delay. If it performs differently than the MA-matched control group, it is classified as exhibited deviance or atypicality. Under the trajectory view, delay corresponds to three types of relationship between the TD trajectory and the disorder trajectory, where both generate reliable linear trajectories: delayed onset, slowed rate, and delayed-onset + slowed rate. These descriptions depend on the significance or nonsignificance of differences in the intercept or gradient of regression lines. Atypicality (deviance, disruption) corresponds to four possibilities, which are discussed in the paragraphs that follow.

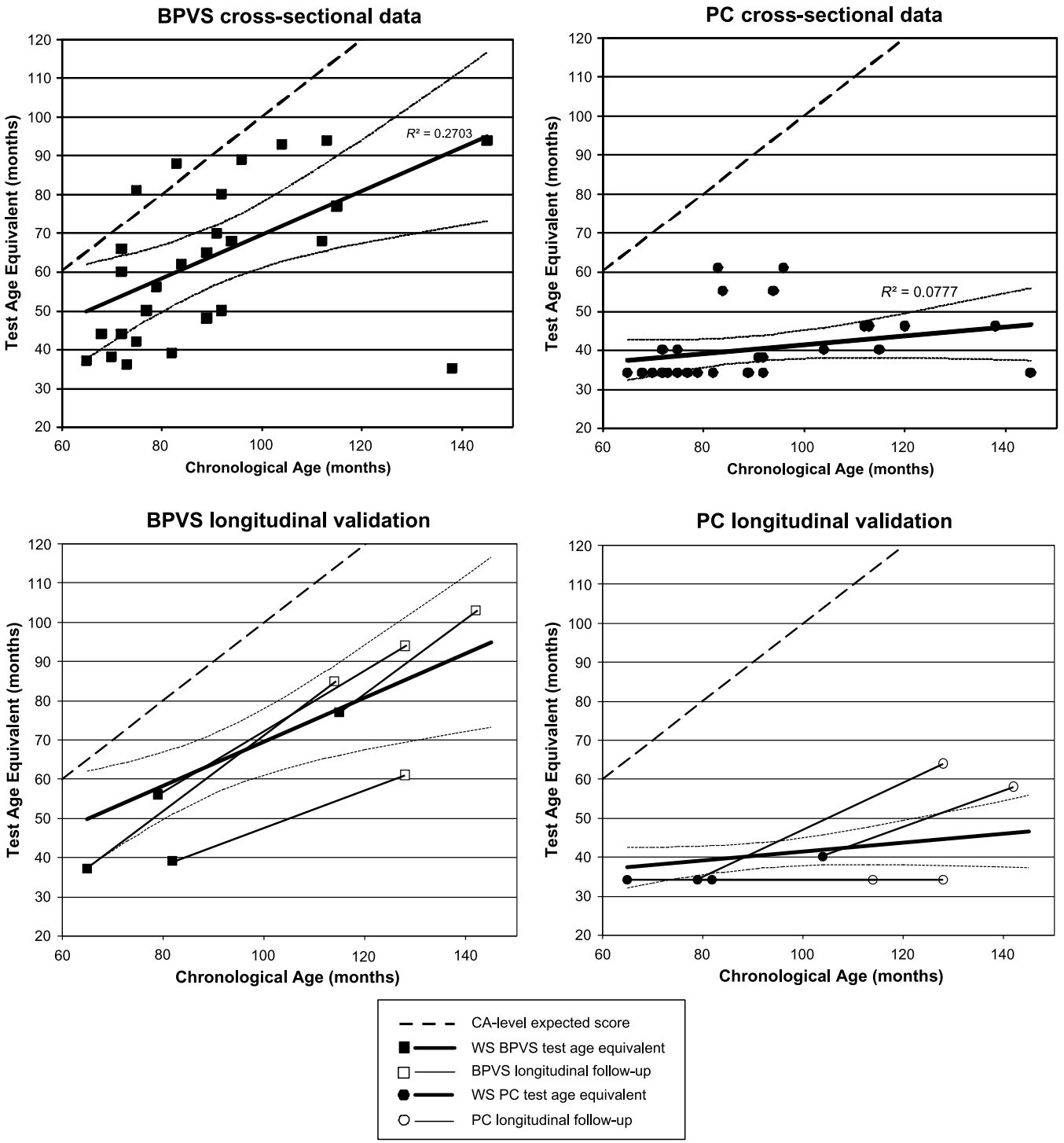
Possibility 1. Although a reliable linear trajectory exists for the TD group, a nonlinear function is a better fit for the disorder group, or there is no reliable trajectory for the disorder group. In the latter case, we distinguished between a zero trajectory and no systematic relationship. Particularly in longitudinal studies, a zero trajectory on an ability assessed with a sensitive measure implies a system that has reached its limit in undergoing ontogenetic change.

Possibility 2. Neither CA nor any theoretically relevant MA measure predicts performance in the disorder group, whereas it does in the TD group.

Possibility 3. A (potentially theoretically unexpected) measure of MA predicts performance in the disorder group but not in the TD group. As we have seen, in cross-sectional designs, one must ensure that the novel association is not an artifact of variations in severity present in the disorder group but not the TD group.

Possibility 4. The same measures of MA predict performance to different extents in the typical and disorder groups.

Figure 10. Comparison of test age scores for 28 children with WS on the British Picture Vocabulary Scale, Second Edition (BPVS-II; Dunn et al., 1997) and the Pattern Construction (PC) subtest of the British Ability Scales, Second Edition (BAS-II; Elliott et al., 1996) plotted against CA. The bold lines show best-fit linear trajectories, along with 95% confidence intervals (thin lines). In the lower panels, unfilled symbols show longitudinal follow-up scores for 4 of the children, within thin lines illustrating individual longitudinal trajectories.



Note that the ascriptions of atypicality based on unexpected developmental relations appeal to an implicit mechanistic account where the cognitive system is taken to develop in integrated blocks or domains (e.g., verbal, nonverbal, spatial). The lack of an expected MA predictor might indicate the absence of the integrated block in the disorder, whereas the presence of an unexpected MA predictor might indicate atypical blocks or developmental contingencies. Under our definitions, one case is problematic. This is when *every* component of a domain appears to be delayed in its development (i.e., is predicted by MA on a measure taken to index the domain), but the delay is different across domains (e.g., verbal, nonverbal, spatial). In this case, the atypicality lies in the differential delay, marked solely by anomalous developmental relations.

In each of these cases of atypicality, the markers of "qualitative" difference rely on (sometimes arbitrary) quantitative cutoffs—that is, that a nonlinear function gives a better fit than a linear function or that the relationship between a predictor and performance is significantly different between groups. Importantly, the only nonquantitative way to identify deviance over delay in a disorder at the level of mechanism relies on the intuition of the experimenter in classifying errors. If a disorder group produces errors that are deemed qualitatively different based on the researcher's experience, a marker of atypicality is claimed (for examples of using errors to test for atypical mechanism, see, e.g., Scerif et al., 2004; Capirci, Sabbadini, & Volterra, 1996; Karmiloff-Smith et al., 1997; Phillips, Jarrold, Baddeley, Grant, & Karmiloff-Smith, 2004; Thomas et al., 2006). The extent to which cases of atypical development occur in developmental disorders remains controversial. For example, Leonard (1998, p. 35) argued that there are very few examples in the literature on SLI that warrant the description of a deviant developmental pattern (for similar arguments, see also Clahsen & Temple, 2003; Tager-Flusberg, 2000; see Thomas, Purser, & Richardson, in press, for discussion).

Discussion

We began by considering two contrasting theoretical positions on the origins of learning disability, the developmental and difference stances (Bennett-Gates & Zigler, 1998) and a more recent instantiation of this distinction in classifying individual cognitive abilities as delayed or deviant/atypical in developmental disorders (Leonard, 1998). For language disorders, Rice et al. (2005) have argued that "the contrast between delayed versus deviant aspects of language acquisition shows considerable promise in providing an overarching perspective on the ways in which language impairments can be manifest" (p. 21). The idea of delay depends on identifying resemblances between the cognitive abilities of a disorder group and those of a younger TD group. In the course of this article, we have argued that the opportunity to find these resemblances depends, to some extent, on the experimental methodology being employed. The use of developmental trajectories provides more ways in which similarities can occur between a disorder group and younger TD controls than the use of matched control groups. A richer descriptive vocabulary for characterizing the ways in which TD can be deflected can only be helpful in seeking causal explanations for the impairments we observe in different disorders.

More widely, we view the strengths and limitations of the trajectory approach as follows. First, trajectories encourage researchers to place the developmental process at the heart of explanations of developmental deficits (Karmiloff-Smith, 1998). Although a methodology brings with it no necessary theoretical commitment, the requirement to derive a function characterizing behavioral change over time focuses research in a way that can sometimes be lost when age (and, therefore, time) is factored out of the design, as is the case in matching. Second, trajectories allow for flexible matching, offering multiple comparisons between the disorder group and a task-specific typical developmental trajectory. Trajectories constructed against CA provide a more theory-neutral characterization of a disorder. Trajectories constructed against MA measures or other experimental tasks allow the researcher to explore developmental relations between abilities. Third, trajectories can be descriptively powerful, as illustrated by the way that they discriminated between different forms of developmental delay. Fourth, although the easiest trajectories to construct are cross-sectional, validation by longitudinal follow-up provides an efficient and productive design.

Fifth, on the downside, the trajectories method relies on testing a wide age range of participants and the availability of tests with sensitivity across that range. Where the behavior of interest is only found in a narrow age range, or tests have limited sensitivity, trajectories are not an optimal design, and matching may be better. Sensitivity is a particular challenge for researchers in the field of language, where early developing abilities such as speech recognition can show ceiling effects. However, fairly complex aspects of language processing can employ relatively simple behavioral responses, thereby making them amenable to use both in young children and in atypical populations. For example, Dick and colleagues (2001) used a comprehension task in which a sentence is presented involving two animals (e.g., "the dog is bitten by the cat"), and the participant is merely required to make a binary response as to which animal is doing the bad action. Sentence construction type can be varied, and performance can be manipulated by perceptual, semantic, and attentional factors to assess

the robustness of processing. This task has been used with participants from 5 to 51 years of age in the typical population, in adult aphasics of various kinds, in children with early focal lesions, and in SLI populations (Dick et al., 2001; Dick, Wulfeck, Krupa-Kwiatkowski, & Bates, 2004; Leech, Aydelott, Symons, Carnevale, & Dick, 2007). We would like to take this opportunity to encourage researchers to develop more measures of language processing with a similar range of sensitivity. It is essential that empirical methods keep pace with the advances being made in research designs for the study of atypical development.

Finally, related to the issue of sensitivity is the question of conceptual cohesion: Tests must have a high probability of indexing the same process across a wide age range. The study of behavior across a wide range opens the trajectory approach to the criticism that there is no guarantee that behavior on the same test is being driven by the same process at different ages. Indeed, there may even be a difference between the TD and disorder groups on the processes responsible for performance at different ages. This is an intrinsic problem in studying development and one that motivates an appeal to multiple converging sources of evidence, such as those provided by developmental cognitive neuroscience.

Acknowledgment

This research was supported by U.K. Medical Research Council Grant G0300188 and European Commission (EC) Grant 0209088 (New and Emerging Science and Technology [NEST]) to Michael Thomas.

References

Annaz, D. (2006). *The development of visuospatial processing in children with autism, Down syndrome, and Williams syndrome*. Unpublished PhD thesis, University of London.

Annaz, D., Karmiloff-Smith, A., Johnson, M. H., & Thomas, M. S. C. (2009). *A cross-syndrome study of the development of holistic face recognition in children with autism, Down syndrome, and Williams syndrome*. *Journal of Experimental Child Psychology* [E-pub ahead of print].

Annaz, D., Karmiloff-Smith, A., & Thomas, M. S. C. (2008). The importance of tracing developmental trajectories for clinical child neuropsychology. In J. Reed & J. Warner-Rogers (Eds.), *Child neuropsychology: Concepts, theory and practice* (pp. 7–18). Oxford, United Kingdom: Wiley-Blackwell.

Ansari, D., Donlan, C., & Karmiloff-Smith, A. (2007). Typical and atypical development of visual estimation abilities. *Cortex, 43,* 758–768.

Baddeley, A. D., Emslie, H., & Nimmo-Smith, I. (1994). The Doors and People Test: A test of visual and verbal recall and recognition. Bury-St-Edmunds, United Kingdom: Thames Valley Test Co.

Bennett-Gates, D., & Zigler, E. (1998). Resolving the developmental-difference debate: An evaluation of the triarchic and systems theory models. In J. A. Burack, R. M. Hodapp, & E. Zigler (Eds.), *Handbook of mental retardation and development* (pp. 115–131). Cambridge, England: Cambridge University Press.

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 109–127). Oxford, England: Oxford University Press.

Bishop, D. V. M. (1997). Cognitive neuropsychology and developmental disorders: Uncomfortable bedfellows. *Quarterly Journal of Experimental Psychology, 50A,* 899–923.

Brock, J. (2007). Language abilities in Williams syndrome: A critical review. *Development and Psychopathology, 19,* 97–127.

Brock, J., & Jarrold, C. (2004). Language influences on verbal short-term memory performance in Down syndrome: Item and order recognition. *Journal of Speech, Language, and Hearing Research, 47,* 1334–1346.

Brock, J., & Jarrold, C. (2005). Serial order reconstruction in Down syndrome: Evidence for a selective deficit in verbal short-term memory. *Journal of Child Psychology and Psychiatry, 46,* 304–316.

Brock, J., Jarrold, C., Farran, E. K., Laws, G., & Riby, D. M. (2007). Do children with Williams syndrome really have good vocabulary knowledge? Methods for comparing cognitive and linguistic abilities in developmental disorders. *Clinical Linguistics & Phonetics, 21,* 673–688.

Bromberg, H., Ullman, M., Coppola, M., Marcus, G., Kelley, K., & Levine, K. (1994, January). *A dissociation of lexical memory and grammar in Williams Syndrome: Evidence from inflectional morphology.* Paper presented at the Sixth International Professional Conference of the Williams Syndrome Association, San Diego, CA.

Capirci, O., Sabbadini, L., & Volterra, V. (1996). Language development in Williams syndrome: A case study. *Cognitive Neuropsychology, 13,* 1017–1039.

Chapman, R. S., Hesketh, L. J., & Kistler, D. J. (2002). Predicting longitudinal change in language production and comprehension in individuals with Down syndrome: Hierarchical linear modeling. *Journal of Speech, Language, and Hearing Research, 45,* 902–915.

Clahsen, H., & Almazan, M. (1998). Syntax and morphology in Williams syndrome. *Cognition, 68,* 167–198.

Clahsen, H., & Temple, C. (2003). Words and rules in Williams syndrome. In Y. Levy & J. Schaeffer (Eds.), *Towards a definition of specific language impairment in children* (pp. 323–352). Mahwah, NJ: Erlbaum.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361.

Cornish, K., Scerif, G., & Karmiloff-Smith, A. (2007). Tracing syndrome-specific trajectories of attention over the lifespan. *Cortex, 43,* 672–685.

Curran, P. J., & Hussong, A. M. (2002). Structural equation modeling of repeated measures data. In D. Moskowitz & S. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications* (pp. 59–86). Mahwah, NJ: Erlbaum.

Dick, F., Bates, E., Wulfeck, B., Aydelott, J., Dronkers, N., & Gernsbacher, M. (2001). Language deficits, localization,

and grammar: Evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review, 108,* 759–788.

Dick, F., Wulfeck, B., Krupa-Kwiatkowski, M., & Bates, E. (2004). The development of complex sentence interpretation in typically developing children compared with children with specific language impairments or early unilateral focal lesions. *Developmental Science, 7,* 360–377.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale, Second Edition (BPVS-II)*. Windsor, Berks, United Kingdom: nferNelson.

Elliott, C. D., Smith, P., & McCulloch, K. (1996). *British Ability Scales, Second Edition (BAS-II)*. Windsor, Berks, United Kingdom: nferNelson.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Grant, J., Valian, V., & Karmiloff-Smith, A. (2002). A study of relative clauses in Williams syndrome. *Journal of Child Language, 29,* 403–416.

Hodapp, R. M., Burack, J. A., & Zigler, E. (Eds.). (1990). *Issues in the developmental approach to mental retardation*. New York: Cambridge University Press.

Hodapp, R. M., & Zigler, E. (1990). Applying the developmental perspective to individuals with Down syndrome. In D. Cicchetti & M. Beeghly (Eds.), *Children with Down syndrome: A developmental perspective* (pp. 1–28). New York: Cambridge University Press.

Jarrold, C., Baddeley, A. D., Hewes, A. K., & Phillips, C. (2001). A longitudinal assessment of diverging verbal and non-verbal abilities in the Williams syndrome phenotype. *Cortex, 37,* 423–431.

Jarrold, C., Baddeley, A. D., & Phillips, C. (2007). Long-term memory for verbal and visual information in Down syndrome and Williams syndrome: Performance on the Doors and People Test. *Cortex, 43,* 233–247.

Jarrold, C., & Brock, J. (2004). To match or not to match? Methodological issues in autism-related research. *Journal of Autism and Developmental Disorders, 34,* 81–86.

Jarrold, C., Cowan, N., Hewes, A. K., & Riby, D. M. (2004). Speech timing and verbal short-term memory: Evidence for contrasting deficits in Down syndrome and Williams syndrome. *Journal of Memory and Language, 51,* 365–380.

Karmiloff-Smith, A. (1997). Crucial differences between developmental cognitive neuroscience and adult neuropsychology. *Developmental Neuropsychology, 13,* 513–524.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences, 2,* 389–398.

Karmiloff-Smith, A., Grant, J., Berthoud, I., Davies, M., Howlin, P., & Udwin, O. (1997). Language and Williams syndrome: How intact is "intact"? *Child Development, 68,* 246–262.

Karmiloff-Smith, A., Scerif, G., & Ansari, D. (2003). Double dissociations in developmental disorders: Theoretically misconceived, empirically dubious. *Cortex, 39,* 161–163.

Karmiloff-Smith, A., Thomas, M. S. C., Annaz, D., Humphreys, K., Ewing, S., Grice, S., et al. (2004). Exploring the Williams syndrome face processing debate: The importance of building developmental trajectories. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45,* 1258–1274.

Karmiloff-Smith, A., Tyler, L. K., Voice, K., Sims, K., Udwin, O., Howlin, P., & Davies, M. (1998). Linguistic dissociations in Williams syndrome: Evaluating receptive syntax in on-line and off-line tasks. *Neuropsychologia, 36,* 343–351.

Klein, B. P., & Mervis, C. B. (1999). Cognitive strengths and weaknesses of 9- and 10-year-olds with Williams syndrome or Down syndrome. *Developmental Neuropsychology, 16,* 177–196.

Leech, R., Aydelott, J., Symons, G., Carnevale, J., & Dick, F. (2007). The development of sentence interpretation: Effects of perceptual, attentional, and semantic interference. *Developmental Science, 10,* 794–813.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Mareschal, D., Johnson, M., Sirios, S., Spratling, M., Thomas, M. S. C., & Westermann, G. (2007). *Neuroconstructivism, Volume 1: How the brain constructs cognition*. Oxford, England: Oxford University Press.

McCauley, R. J., & Swisher, L. (1984). Use and misuse of norm-referenced tests in clinical assessment: A hypothetical case. *Journal of Speech and Hearing Disorders, 49,* 333–348.

Mervis, C. B., & Robinson, B. F. (2003). Methodological issues in cross-group comparisons of language and cognitive development. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations: Towards a definition of specific language impairment* (pp. 233–258). Mahwah, NJ: Erlbaum.

Mervis, C. B., Robinson, B. F., Bertrand, J., Morris, C. A., Klein-Tasman, B. P., & Armstrong, S. C. (2000). The Williams syndrome cognitive profile. *Brain & Cognition, 44,* 604–628.

Meyer-Lindenberg, A., Mervis, C. B., & Berman, K. F. (2006). Neural mechanisms in Williams syndrome: A unique window to genetic influences on social cognition and behavior. *Nature Reviews Neuroscience, 7,* 380–393.

Motulsky, H., & Christopoulos, A. (2004). *Fitting models to biological data using linear and nonlinear regression: A practical guide to curve fitting*. Oxford, England: Oxford University Press.

Paterson, S. J., Brown, J. H., Gsödl, M. K., Johnson, M. H., & Karmiloff-Smith, A. (1999, December 17). Cognitive modularity and genetic disorders. *Science, 286,* 2355–2358.

Peugh, J. L., & Enders, C. K. (2005). Using the SPSS mixed procedure to fit cross-sectional and longitudinal methods. *Educational and Psychological Measurement, 65,* 717–741.

Phillips, C. E., Jarrold, C., Baddeley, A., Grant, J., & Karmiloff-Smith, A. (2004). Comprehension of spatial language terms in Williams syndrome: Evidence for an interaction between domains of strength and weakness. *Cortex, 40,* 85–101.

Rice, M. L. (2004). Growth models of developmental language disorders. In M. L. Rice & S. F. Waren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 207–240). Mahwah, NJ: Erlbaum.

Rice, M. L., Tomblin, J. B., Hoffman, L., Richman, W. A., & Marquis, J. (2004). Grammatical tense deficits in

children with SLI and nonspecific language impairment: Relationships with nonverbal IQ over time. *Journal of Speech, Language, and Hearing Research, 47,* 816–834.

Rice, M. L., Warren, S. F., & Betz, S. (2005). Language symptoms of developmental language disorders: An overview of autism, Down syndrome, fragile X, specific language impairment, and Williams syndrome. *Applied Psycholinguistics, 26,* 7–27.

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1412–1431.

Rossen, M., Klima, E. S., Bellugi, U., Bihrle, A., & Jones, W. (1996). Interaction between language and cognition: Evidence from Williams syndrome. In J. H. Beitchman, N. Cohen, M. Konstantareas, & R. Tannock (Eds.), *Language learning and behavior* (pp. 367–392). New York: Cambridge University Press.

Scerif, G., Cornish, K., Wilding, J., Driver, J., & Karmiloff-Smith, A. (2004). Visual selective attention in typically developing toddlers and toddlers with fragile X and Williams syndrome. *Developmental Science, 7,* 118–130.

Scerif, G., Karmiloff-Smith, A., Campos, R., Elsabbagh, M., Driver, J., & Cornish, K. (2005). To look or not to look? Typical and atypical development of oculomotor control. *Journal of Cognitive Neuroscience, 4,* 591–604.

Schopler, E., Reichler, R., & Rochen, B. (1993). *The Childhood Autism Rating Scale*. Los Angeles: Western Psychological Services.

Singer Harris, N. G., Bellugi, U., Bates, E., Jones, W., & Rossen, M. (1997). Contrasting profiles of language development in children with Williams and Down syndromes. *Developmental Neuropsychology, 13,* 345–370.

Strauss, M. E. (2001). Demonstrating specific cognitive deficits: A psychometric perspective. *Journal of Abnormal Psychology, 110,* 6–14.

Tager-Flusberg, H. (2000). Differences between neurodevelopmental disorders and acquired lesions. *Developmental Science, 3,* 33–34.

Tager-Flusberg, H., Plesa-Skwerer, D., Faja, S., & Joseph, R. M. (2003). People with Williams syndrome process faces holistically. *Cognition, 89,* 11–24.

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

Thomas, M. S. C., Dockrell, J. E., Messer, D., Parmigiani, C., Ansari, D., & Karmiloff-Smith, A. (2006). Speeded naming, frequency, and the development of the lexicon in Williams syndrome. *Language and Cognitive Processes, 21,* 721–759.

Thomas, M. S. C., Grant, J., Gsödl, M., Laing, E., Barham, Z., Lakusta, L., et al. (2001). Past tense formation in Williams syndrome. *Language and Cognitive Processes, 16,* 143–176.

Thomas, M. S. C., & Karmiloff-Smith, A. (2002). Are developmental disorders like cases of adult brain damage? Implications from connectionist modeling. *Behavioral and Brain Sciences, 25,* 727–780.

Thomas, M. S. C., & Karmiloff-Smith, A. (2003). Modeling language acquisition in atypical phenotypes. *Psychological Review, 110,* 647–682.

Thomas, M. S. C., Purser, H. R. M., & Richardson, F. M. (in press). Modularity and developmental disorders. In P. D. Zelazo (Ed.), Oxford handbook of developmental psychology. Oxford, United Kingdom: Oxford University Press.

Torppa, M., Poikkeus, A., Laakso, M., Tolvanen, A., Leskinen, E., Leppanen, P., Puolakanaho, A., & Lyytinen, H. (2007). Modeling the early paths of phonological awareness and factors supporting its development in children with and without familial risk of dyslexia. *Scientific Studies of Reading, 11,* 73–103.

van der Lely, H. K. J., & Ullman, M. T. (2001). Past tense morphology in specifically language impaired and normally developing children. *Language and Cognitive Processes, 16,* 177–217.

van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review, 98,* 3–53.

Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology, 10,* 395–426.

Yule, W. (1978). Research methodology: What are the "correct controls"? In M. Rutter & E. Schopler (Eds.), *Autism: A reappraisal of concepts and treatment* (pp. 155–162). New York: Plenum.

Zukowski, A. (2001). *Uncovering grammatical competence in children with Williams syndrome*. Unpublished doctoral thesis, Boston University.

Received July 6, 2007

Revision received January 7, 2008

Accepted August 7, 2008

DOI: 10.1044/1092-4388(2009/07-0144)

Contact author: Michael Thomas, Developmental Neurocognition Laboratoy, School of Psychology, Birkbeck College, University of London, Malet Street, Bloomsbury, London WC1E 7HX, United Kingdom.
E-mail: m.thomas@bbk.ac.uk.