Psychological Review
2003, Vol. 110, No. 4, 647–682

0033-295X/03/$12.00 DOI: 10.1037/0033-295X.110.4.647

Modeling Language Acquisition in Atypical Phenotypes

Michael S. C. Thomas and Annette Karmiloff-Smith
Institute of Child Health

An increasing number of connectionist models have been proposed to explain behavioral deficits in developmental disorders. These simulations motivate serious consideration of the theoretical implications of the claim that a developmental disorder fits within the parameter space of a particular computational model of normal development. The authors examine these issues in depth with respect to a series of new simulations investigating past-tense formation in Williams syndrome. This syndrome and the past-tense domain are highly relevant because both have been used to make strong theoretical claims about the processes underlying normal language acquisition. The authors conclude that computational models have great potential to advance psychologists' understanding of developmental deficits because they focus on the developmental process itself as a pivotal causal factor in producing atypical phenotypic outcomes.

Computational models have become an increasingly prevalent tool for investigating mechanisms of change within cognitive development (e.g., Simon & Halford, 1995). Much of this research has used connectionist learning systems (i.e., computer models loosely based on principles of neural information processing) to construct cognitive-level explanations of behavior (Elman et al., 1996; Mareschal & Thomas, 2001). Such models have offered a way to explore self-organization in development, the process whereby structure emerges in a representational system in response to the system's dynamic interactions with its environment. Self-organization is guided by constraints or boundary conditions built into the initial state of the system, and computational models have permitted researchers to investigate how different system constraints interact with an environment to generate observed behaviors.

In addition to studying normal development, these models have provided a means of exploring how deviations in self-organization, due to a shift in initial constraints, can result in the emergence of atypical behaviors such as those found in developmental disorders (Mareschal & Thomas, 2001; Oliver, Johnson, Karmiloff-Smith, & Pennington, 2000; Thomas & Karmiloff-Smith, 2002b).

Although, in principle, any type of developmental computational model can be applied to the study of developmental disorders, thus far most models have appeared within the connectionist paradigm. Developmental connectionist models contain a number of initial parameter and design decisions made by the modeler prior to the learning process. These decisions include the initial architecture of the model, the activation dynamics of the processing units, the choice of input–output representations, the type of learning algorithm, and the nature of the training set. Increasing numbers of models have been put forward as offering explanations of deficits in developmental disorders on the basis of alterations to these initial constraints. During training, such models can exhibit an atypical trajectory of development with behavioral impairments emerging in their end states.

Three domains—dyslexia, autism, and specific language impairment—serve to illustrate this approach. Take, for example, *reading*. Developmental phonological dyslexia has been explained via manipulations to initial phonological and orthographic representations of a connectionist model. Alternatively, researchers have proposed the use of a two-layer network or a reduction in hidden unit numbers in the initial architecture, or alterations to the learning algorithm and/or the architecture of a subsystem that learns the phonological forms of words (Brown, 1997; Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998a). Several proposals also exist for initial manipulations that might capture developmental surface dyslexia. These proposals include a reduction in the number of hidden units, a less efficient learning algorithm, less training, and a slower learning rate (Bullinaria, 1997; Harm & Seidenberg, 1999; Plaut et al., 1996; Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998b). In *autism*, categorization deficits have been explained in terms of network architectures that have too few or too many hidden units, noise vectors added to the input (Cohen, 1994, 1998), or self-organizing feature maps with exaggerated levels of lateral inhibition (Gustafsson, 1997; see for discussion, Mareschal & Thomas, 2001; Thomas, 2000). In *specific language impairment* (SLI), deficits in inflectional morphology have been explained in terms of a network with initially degraded phonological representations (Hoeffner & McClelland, 1993; Joanisse, 2000).

Michael S. C. Thomas and Annette Karmiloff-Smith, Neurocognitive Development Unit, Institute of Child Health, London.

This work was funded by Medical Research Council (MRC) Project Grant G9809880 and MRC Programme Grant G9715642 to Annette Karmiloff-Smith.

We thank Denis Mareschal and Birkbeck College; Kim Plunkett and the McDonnell-Pew Centre for Cognitive Neuroscience for the use of time on their computers; Ulrike Hahn and David Green for comments on a draft of the manuscript; John Bullinaria for suggesting the use of plasticity and in particular the sigmoid prime parameter; Deborah Unthank for help and advice in developing the phonological coding schemes; Neil Forrester for computational advice; and Jay McClelland, Steven Pinker, Gary Marcus, Gert Westermann, and Denis Mareschal for helpful discussions.

Correspondence concerning this article should be addressed to Michael S. C. Thomas, who is now at the School of Psychology, Birkbeck College, University of London, Malet Street, London WC1E 7HX, United Kingdom. E-mail: m.thomas@psychology.bbk.ac.uk

This conception of developmental disorders has major advantages but also potential limitations. One advantage is that developmental computational models allow a proper consideration of the crucial role of the developmental process itself in producing behavioral deficits, in contrast to a widespread view that developmental disorders can be explained within a static framework as the direct analogue of acquired disorders (Thomas & Karmiloff-Smith, 2002a). One potential limitation arises from the claim that disorders fit within the parameter space of particular computational implementations. Such a claim raises a number of contentious issues, including the relation of simulation to explanation, the validity of any given implementation, and the flexibility of that model in capturing various patterns of developmental data. In the following paragraphs, we consider these points in more detail.

To understand the benefit of using connectionist models in studying developmental disorders, we must first review the explanatory framework within which such disorders are typically conceived. The field of developmental cognitive neuropsychology began as an extension of the adult cognitive neuropsychological model to data from children with neuropsychological disorders. The initial explanatory framework, therefore, assumed a static modular structure to the cognitive system and sought to characterize developmental disorders in terms of the atypical development of one or more components, assumed from theories of normal cognitive functioning. This extension is illustrated by an emphasis on the search for double dissociations of cognitive functions between different developmental disorders (Temple, 1997), a pattern of empirical data with particular significance in the adult framework because it is taken as a strong indication of damage to independent cognitive components.

Because behavioral impairments in developmental disorders are usually identified in children and adults when many of the developmental processes are close to their end state, such impairments are often compared against a static description of the functional structure of the normal cognitive system. This sometimes encourages analogies to be drawn between developmental and acquired deficits. In such cases, there is an assumption that a deficit in behavior at the end of development (i.e., the outcome of a developmental process) can be mapped one-to-one onto a deficit in one or more cognitive mechanisms caused by damage to an adult system, while in both cases the rest of the system is intact and functioning normally. Baron-Cohen (1998) summarized this view: "I suggest that the study of mental retardation would profit from the application of the framework of cognitive neuropsychology (e.g., McCarthy & Warrington, 1990; Shallice, 1988). In cognitive neuropsychology, one key question running through the investigator's mind is 'Is this process or mechanism intact or impaired in this person?'" (p. 335).

The advantage of interpreting acquired and developmental disorders within the same framework is the possibility of accessing two sources of complementary evidence that may converge to reveal the structure of the cognitive system. Thus, Temple (1997) discussed a range of behavioral impairments for which acquired and developmental analogues can be found (see Thomas & Karmiloff-Smith, 2002a, for discussion). The two sources of information tell us different things. Acquired deficits can reveal the structure of the adult system, whereas truly selective developmental deficits can demonstrate components that develop independently. Furthermore, when developmental disorders have a genetic basis, perhaps truly selective behavioral deficits (if there are any) may be evidence of innate modular structure in the cognitive system, in this case selectively damaged by a genetic anomaly.

The difficulty with interpreting developmental deficits within a static modular framework is that such accounts exclude the developmental process as a causal factor in the disorder (see Karmiloff-Smith, 1997, 1998, for discussion). This is particularly problematic when the modular structure itself appears to be the product of a developmental process. A growing number of studies show how both neural localization and neural specialization for biologically important functions such as species recognition and language take place gradually across development (M. Johnson, 1999; Neville, 1991). To achieve a selective high-level deficit against a background of normal functioning in a developmental system would require very strong and perhaps unrealistic assumptions about the constraints that guide the developmental process as well as limitations to the extent that compensation can overcome early deficits (Thomas & Karmiloff-Smith, 2002a).

Because innate modularity of high-level functions does not appear to be a viable assumption (see below), selective high-level developmental impairments would then require a picture in which specialized processing components could emerge independently of each other during development (i.e., sufficient independence that early deviations in one mechanism would not affect the development of others). However, Bishop (1997) has argued that interactivity between systems, rather than independence, is the hallmark of early development. In addition, any compensation that developmental plasticity permits is likely to lead to knock-on effects in other domains, in which areas attempting to compensate for malfunctioning systems themselves experience a reduction in efficiency in carrying out their normal functions (see Anderson, Northam, Hendy, & Wrennall, 2001, for discussion).

The hope that genetic developmental disorders can provide evidence of innate modular structure is undermined by an absence of direct links between genes and particular high-level cognitive structures. Currently, there are no known genes that serve the function of coding directly for specific high-level cognitive structures and, consequently, for domain-specific developmental outcomes. Indeed, current knowledge suggests that genetic effects in brain development are generally both widespread and graded, and when they occur in more restricted areas, these do not match up with subsequent regions of specialization for higher cognitive functions, with the possible exception of primary sensory and motor systems (Karmiloff-Smith, 1998; Karmiloff-Smith, Scerif, & Thomas, 2002; Kingsbury & Finlay, 2001).

The alternative to viewing developmental impairments as if they were high-level lesions to a static system is to view them as the outcome of initial differences in the lower level constraints under which the cognitive system develops (i.e., the high-level deficits are an outcome of development itself; Elman et al., 1996; Karmiloff-Smith, 1998; Oliver et al., 2000). When genetic damage leads to high-level anomalies in a developmental disorder, differences are likely to lie in the initial low-level neurocomputational properties of the brain, such as local connectivity or the firing properties of neurons, rather than in selective deficits to high-level cognitive components. Different, initial low-level constraints lead to alternative developmental trajectories, which in turn generate a particular profile of high-level cognitive abilities. This perspective has implications for the types of data that are collected in charac-

terizing developmental disorders. An approach that predicts widespread atypicalities across cognitive domains with more serious and less serious behavioral consequences will generate a different research agenda to one that simply searches for selective deficits against a background of normal function, an issue we consider in more detail elsewhere (Thomas & Karmiloff-Smith, 2002a).

Connectionist models of development are ideally suited for exploring this latter, dynamic view of developmental disorders, because their final behavior is a product of initial (lower level) network constraints and a subsequent developmental process. Alterations in the initial network constraints can cause deficits in performance at the end of training, as well as differences in the stages through which it passes. Models offer the particular advantage of allowing a detailed consideration of the relation between initial constraints and trajectories of development in complex learning systems. Such relationships are hard to anticipate without the use of modeling.

Despite the gains that computational accounts of developmental disorders may offer in their emphasis on the process of development itself as a cause, such accounts are potentially undermined by the limitations of computational modeling. In each of the examples we have introduced (dyslexia, autism, SLI), the explanation of disordered performance amounted to the claim that atypical performance falls within the parameter space of a particular computational model. Yet a claim of this sort raises a number of potential objections. Some of these are specific to the particular model: How does one define (and justify) the parameter set for a normal model in a given domain—the precondition for simulating atypical development? What is the justification for manipulating a particular parameter to fit the disordered data (e.g., changing the number of hidden units in a network)? When psychological data motivate the manipulation of the parameter, is this parameter the only way to implement the deficit suggested by the psychological data? When a parameter manipulation (such as number of hidden units) fits the group data of a disordered population, does this parameter have sufficient scope to cover the full range of individual variation shown by the disorder (e.g., from failure to arrested development to delayed success)? And when one parameter manipulation fits the disordered data, how unique is this finding—how do we know that there are not many parameter manipulations within the model that would also fit the data?

Other objections are more general. If a model happens to fit both the normal and disordered data, how can we guarantee that our chosen model is the right one, with the right number of parameters? For example, connectionist models of reading show a fair degree of variation in their exact design: How can we be sure that a successful manipulation to one model holds for all other models of the domain? In other words, to what extent can we generalize the claims made from any given model?

Despite the increasing emergence of connectionist models of developmental disorders, objections such as these have rarely been given due consideration. If atypical models are to realize their potential, such objections must be evaluated carefully. In this article, our aim is to begin this task. Our starting point is to introduce a concrete example around which we can focus the theoretical discussion, with a target developmental disorder and a target behavioral deficit. The target disorder is Williams syndrome (WS), and the target domain is language development, in particular past-tense acquisition. Several reasons motivate this choice.

First, the domain of past tense offers an excellent example of how researchers have formulated explanations of deficits in developmental disorders based on direct analogies to selective high-level deficits in a static system, including the application of double dissociation methodology to motivate the postulation of independent processing mechanisms. Indeed, past tense offers an example of the use of genetic developmental disorders to bolster claims about innate high-level structure in the language system. Modeling work in this domain may clarify whether such claims are necessary when one adopts a more developmental perspective.

Second, WS is important because deficits in the language of individuals with this disorder have been used to make strong theoretical claims about the nature of typical language development. In constructing our model, we identify several hypotheses concerning the overall cause of atypical language development in WS. Particular claims have been made about past-tense deficits in WS, and modeling work permits us to evaluate whether each hypothesis is sufficient to capture WS past-tense data in a developmental model.

Third, the modeling of atypical past-tense acquisition is made easier by the existence of a body of work that has used connectionist models to simulate typical development in past tense formation. This is important because, before one undertakes a consideration of atypical development from a computational perspective, one must begin with a baseline model of typical development.

Fourth, despite the existence of fairly good connectionist implementations of past-tense acquisition, there is nevertheless a competing theoretical account in this domain (albeit one that is not sufficiently specified to allow computational implementation). The existence of two dominant theories drives a consideration of the generality of the findings of one particular connectionist simulation to other models within the field.

We start, then, with an examination of the way in which developmental disorders have been used to shed light on the structure of the normal past-tense system. We then consider in detail the evidence on inflectional morphology in WS and identify several distinct hypotheses on the wider causes of atypical language development in this syndrome. At this point, we turn to connectionist modeling, first outlining a baseline or normal model and then describing the parameter manipulations that may allow us to simulate a set of target data from a detailed study on past-tense formation in WS. Finally, we return to consider the general use of developmental computational models for the study of developmental disorders.

The English Past Tense and Developmental Disorders

The English past tense is characterized by a predominant regularity in which the majority of verbs form their past tense by the addition of one of three allomorphs of the *–ed* suffix to the base stem (e.g., walk/walked, end/ended, chase/chased). However, there is a small but significant group of verbs that form their past tense in different ways, including changing internal vowels (swim/swam), changing word final consonants (build/built), changing both internal vowels and final consonants (think/thought), an arbitrary relation of stem to past tense (go/went), and verbs that have a past-tense form identical to the stem (hit/hit). These so-called irregular verbs often come in small groups sharing a family re-

semblance (sleep/slept, creep/crept, leap/leapt) and usually have high token frequencies (see Pinker, 1999, for further details).

During the acquisition of the English past tense, children show a characteristic U-shaped developmental profile at different times for individual irregular verbs. Initially they use the correct past tense of a small number of high-frequency regular and irregular verbs. Later, they sometimes produce overregularized past-tense forms for a small fraction of their irregular verbs (e.g., thinked; Marcus et al., 1992) along with other, less frequent errors (Xu & Pinker, 1995). Finally, performance is good on both regular and irregular verbs (Berko, 1958; Ervin, 1964; Kuczaj, 1977).

Currently, two theories compete to explain the cognitive processes underlying past-tense performance: a connectionist theory rooted in implemented computer simulations (e.g., Joanisse & Seidenberg, 1999; Plunkett & Juola, 1999; Plunkett & Marchman, 1991, 1993; Rumelhart & McClelland, 1986) and a more descriptive, dual mechanism theory (e.g., Marcus et al., 1992; Pinker, 1991, 1994, 1999). Both theories take respective performance on regular and irregular past tenses as indexing different things, either different underlying knowledge or different underlying processes. In the connectionist theory, performance on regular verbs indexes reliance on knowledge about phonological regularities, whereas that on irregular verbs indexes reliance on lexical–semantic knowledge. In the dual mechanism theory, performance on regular verbs indexes a dedicated symbolic processing mechanism implementing the regular rule, whereas performance on irregular verbs indexes an associative memory storing information about the past-tense forms of specific verbs. Evidence from developmental disorders has been applied to this latter model, in combination with evidence from acquired disorders.

In terms of acquired disorders, Pinker (1991, 1994, 1999) pointed to evidence from adults with neurodegenerative diseases and acquired aphasia as supporting two separate, qualitatively different, high-level processing mechanisms within the inflectional morphology system. Patients with nonfluent aphasia can be worse at producing and reading regular past-tense forms than exception forms, whereas patients with fluent aphasia can be worse at producing and reading exception forms than regular forms (e.g., Tyler, de Mornay Davies, et al., 2002; Tyler, Randall, & Marslen-Wilson, 2002; Ullman et al., 1997, in press; although see Bird et al., 2003). Similarly, patients with Parkinson's disease can make more errors producing regular and novel *–ed* forms than exception forms, whereas patients with Alzheimer's disease can make more errors producing exception past-tense forms than regular past-tense forms (Ullman, in press; Ullman et al., 1997). Using the logic of adult neuropsychology, these double dissociations are taken as evidence that independent mechanisms are responsible for performance on the two types of verbs.

Within the same framework, Pinker (1991, 1994, 1999) also supported his dual mechanism argument with evidence from individuals who had developmental disorders. In particular, he cited WS as a case in which genetic damage disrupts the functioning of one of the two mechanisms. Damage to the associative memory for exception past tenses leads to a purported selective deficit in irregular past-tense formation. In line with this claim, Clahsen and Almazan (1998) argued that in WS, the computational system for language is selectively spared, whereas the lexical system required for irregular inflection is impaired. More widely, Clahsen and Temple (2003) have claimed that the right way to view the WS language system is in terms of the architecture of a normal system but with selective high-level components that are under- or overdeveloped. This illustrates a clear preference for a static modular framework to explain developmental deficits.

Conversely, Pinker (1991, 1994, 1999) presented SLI as a case of a genetic syndrome offering the opposite pattern to WS, in which the rule-based mechanism is impaired and the associative memory is intact. Like WS, SLI is a developmental disorder with a genetic component (Bishop, North, & Donlan, 1995). In SLI, impairments are found in language in the absence of any apparent cognitive, social, or neurological deficits. Van der Lely and Ullman (2001) found that not only were children with SLI very poor at applying the "add *–ed*" past-tense regularity to novel verbs, but they also showed poor performance on both regular and irregular past tenses, their predominant response being to produce uninflected stems (see also Moore & Johnston, 1993; Ullman & Gopnik, 1999). Indeed, Leonard, Bortolini, Caselli, McGregor, and Sabbadini (1992) and Oetting and Horohov (1997) reported a higher percentage of irregular than regular past-tense usage in SLI. In interpreting their data, Van der Lely and Ullman (2001) suggested that, in the absence of the rule-based mechanism, these children were using their associative memory mechanism to memorize both regular and irregular past tenses. They cited as evidence the unusual presence of frequency effects in the levels of performance of regular verbs. Taking SLI and WS together, Pinker (1999) made the following comment:

> Overall, the genetic double dissociation is striking, suggesting that language is both a specialisation of the brain and that it depends on generative rules that are visible in the ability to compute regular forms. The genes of one group of children [SLI] impair their grammar while sparing their intelligence; the genes of another group of children [WS] impair their intelligence while sparing their grammar. The first group of children rarely generalise the regular pattern; the second group of children generalise it freely. (p. 262)

Three points are illustrated here. First, here is a case of developmental disorders being used in a directly analogous fashion to acquired deficits to support claims for the structure of the adult system. Second, these disorders are analyzed within the same adult cognitive neuropsychological framework, in terms of selective deficits to an otherwise normal system and with the use of the double dissociation methodology. Third, genetic developmental disorders are being used to make claims about innate, high-level structure in the normal language system. Our next task is to summarize the extant data on inflectional morphology in WS.

WS and Past Tense

WS, a rare neurodevelopmental disorder, has been of theoretical interest because it exhibits an uneven cognitive–linguistic profile while being caused by the deletion of only a small number of genes (see Donnai & Karmiloff-Smith, 2000, for review). Overall IQ scores, which typically fall between 50 and 70, mask differences in specific cognitive abilities: Individuals frequently show fairly good verbal abilities with deficient visuospatial abilities. Although people with WS often perform within the normal range on certain standardized tests for face recognition (Bellugi, Wang, & Jernigan, 1994; Udwin & Yule, 1991) and show relatively good performance on theory-of-mind tasks (Karmiloff-Smith, Klima, Bellugi, Grant,

& Baron-Cohen, 1995), they exhibit difficulties in numerical cognition (Karmiloff-Smith et al., 1995) and in problem solving and planning (Bellugi, Marks, Bihrle, & Sabo, 1988).

Investigations of inflectional morphology in WS have focused on several tasks, including past-tense elicitation, plural formation, noun compounding, and comparative formation. However, many such studies have been compromised by the use of very small participant numbers against a background of marked individual variability within the syndrome. Clahsen and Almazan (1998, 2001) reported a selective deficit in irregular past-tense formation compared with mental age (MA)-matched controls in a study involving 4 participants with WS. An equivalent deficit was found in a plural formation task, but in a second pluralization task, no significant irregular deficit was reported.[1]

When Thomas et al. (2001) examined past-tense production in a much larger sample of 18 individuals with WS, no selective irregular verb deficit was found after differences in verbal MA had been controlled for. Although a disparity in irregular verb performance was apparent compared with chronological age matches, this appeared to be the consequence of a combination of delayed language development and the normal lag between regular and irregular past-tense acquisition. The study did reveal a deficit in generalization of past-tense regularities to novel forms in the WS group, as well as an increased semantic effect (of verb imageability) on irregular past-tense performance in the WS group compared with controls.

Zukowski (2001) examined pluralization in a sample of 12 individuals with WS and again failed to find a significant selective deficit for irregular pluralization but also failed to find a reduction in generalization. Zukowski noted that unlike regular plural formation, irregular formation in the WS group improved when individuals were prompted for a further response. Initially unmarked forms (mouse–mouse) were correctly inflected with further prompting (mouse–mice), as if participants with WS knew that an irregular plural existed, thus suppressing regularization (mouse–mouses) but had difficulty in retrieving the irregular form. This pattern was not found in the control group.

Clahsen and Almazan (2001) presented evidence that their 4 participants with WS were not sensitive to constraints that prevent regular plurals appearing in noun–noun compounds (*rats–eater) but allow irregular plurals to appear (mice–eater). They argued that this was theoretically consistent with a selective deficit to irregular inflection. However, once more, Zukowski (2001) failed to replicate this effect with a larger sample.

What then should we take to be the pattern of WS past-tense performance? The most salient feature is developmental delay in inflecting both regular and irregular forms. The largest study (Thomas et al., 2001) suggested this delay was equal across verb types and accompanied by reduced generalization. A reduction in generalization has been reported in other morpho–phonological language tasks with individuals who have WS (e.g., gender agreement in French: Karmiloff-Smith et al., 1997; in Hebrew, coining of correct agent nouns from verb roots, gender agreement on animate nouns, and the well-formedness of derived verb forms: Levy & Hermon, in press). Although Zukowski (2001) did not find this reduction in English pluralization, the English plural paradigm is marked by a quantitatively different distribution, with a much greater proportion of regular to irregular forms than in past tense (see Plunkett & Juola, 1999). In terms of a greater (specific) delay for irregular inflection, the larger studies have failed to find this pattern. Nevertheless, such a difference has been reported in two other small studies, in comparative formation (Clahsen & Temple, 2003) and in German plural formation (Krause & Penke, 2000).

Because the Thomas et al. (2001) study involved the largest participant numbers and the largest stimulus sets, we take this pattern of data to be the target of our modeling of WS past-tense formation. These data are advantageous for this kind of simulation because they compare empirically the relative developmental trajectories of a WS group and a typical control group on a past-tense elicitation task. Such trajectories can be matched against those derived from a developmental computational model. Conversely, given the reports of irregular deficits from some smaller WS studies, we also examine qualitatively what manipulations to the start state of the model could lead to this alternate pattern.

We now consider what initial anomalies in lower level constraints may underlie the differential pattern of past-tense acquisition seen in WS. These constraints determine the manipulations that are applied to the starting state of a computational model of normal development, as described in the Implementing the Hypotheses for Differential Constraints in WS Language Acquisition section, below.

Atypical Constraints in WS Language Development

Language in WS was initially portrayed as developing normally, despite low general cognitive ability. This led to some excitement that the disorder might represent an existence proof of the developmental independence of language and cognition (e.g., Pinker, 1991). However, subsequent research has suggested that in most areas of language, WS performance is more in line with MA controls than chronological age controls, arguing against such a developmental independence. It is certainly the case that when compared with language development in other genetic syndromes like Down's syndrome with equivalent general cognitive abilities, WS language appears much more advanced. However, detailed research has nevertheless revealed atypicalities in WS language at all levels of performance.

For example, examination of precursors to language development in toddlers with WS revealed reduced levels of pointing and impairments in triadic interactions, both important bases for the development of referential language use (Laing et al., 2002). When language appears, it is usually delayed (e.g., in a study by Singer-Harris, Bellugi, Bates, Jones, & Rossen, 1997, the mean delay was 2 years). The vocabulary spurt, when it occurs, does not appear to be associated with markers of maturing semantic knowledge in the same way that it is in typically developing children (Mervis & Bertrand, 1997). There is some suggestion that lexical constraints used in vocabulary acquisition are also atypical (Stevens & Karmiloff-Smith, 1997) as well as some evidence that the normal production–comprehension asymmetry may be reduced in WS (Paterson, 2000). Some data even suggest that children with WS produce more words than they comprehend, as if they were mem-

[1] Bromberg et al. (1994) reported a similar preliminary finding of a pluralization on a sample of 6 individuals with WS in an unpublished conference presentation, but without analyses to demonstrate that differences were significant.

orizing and using phonological forms without a firm grip of their semantic underpinnings (Singer Harris et al., 1997).

Studies have also pointed to difficulties in the acquisition of morphosyntax in WS, particularly in languages with complex morphology. For example, difficulties in gender assignment in WS have been reported in a number of languages (Spanish: Cáceres, Heinze, & Méndez, 1999; French: Karmiloff-Smith et al., 1997; Italian: Volterra, Capirci, Pezzini, Sabbadini, & Vicari, 1996). Some errors found in the acquisition of morphology were qualitatively different from those found in normal development (Capirci, Sabbadini, & Volterra, 1996). Studies of syntax have indicated a greater delay for grammar acquisition than vocabulary acquisition, and although a similar overall pattern of ease and difficulty is seen in the production of different constructs in WS and MA controls, the difficulty is often exaggerated for the WS group (e.g., for relative clauses; see Grant, Valian, & Karmiloff-Smith, 2002; Zukowski, 2001).

Much work has been directed towards investigating lexical semantics in WS, prompted by the sometimes unusual vocabulary that individuals incorporate into their expressive language. The picture here appears to be that atypical vocabulary use is either pragmatic or reflects poorer underlying knowledge. Lexical access operates more slowly in WS but exhibits normal processing dynamics (Thomas, Dockrell, et al., 2002). The final semantic and conceptual representations formed in individuals with WS appear to be shallower, with less abstract information and more perceptually based detail, suggested by studies examining conceptual knowledge in WS (S. Johnson & Carey, 1998) and the development of semantic categories and metaphorical relations (Thomas, van Duuren, Ansari, Parmigiani, & Karmiloff-Smith, 2002). Last, individuals with WS also show pragmatic deficits, with speech content often odd or out of place in a particular social context (Volterra, Capirci, & Caselli, 2001), high levels of clichés and stereotyped phrases (Howlin, Davies, & Udwin, 1998a), and evidence of difficulties in comprehending nonliteral language (Howlin, Davies, & Udwin, 1998b).

From a detailed review of the literature, we identified two types of hypothesis concerning the underlying causes of atypical language development in WS. The first is a conservative hypothesis, in which it is argued that the language we see in WS is merely the product of delayed development combined with low IQ. In this view, aspects of language performance that are specific to WS are indirect effects of other characteristics of the disorder. The second explanation is perhaps better viewed as a cluster of related hypotheses, broadly falling under what we call the semantics–phonology imbalance theory.

The conservative hypothesis can be described as follows: Deficits in syntax and pragmatics in WS are what one might expect at a given level of mental retardation. Language development from the earliest age reflects the interests of a child with WS, specifically a strong desire for social interaction (e.g., Jones et al., 2000). Language is initially used more to mediate these interactions than as a referential tool. Subsequent vocabulary development reflects the special interests of the child who has some degree of mental retardation, with unusual word usage strategically used to gain attention and mediate social interaction. Deficits that do exist in vocabulary reflect other nonlinguistic aspects of WS. For example, children's visuospatial processing deficit leads to problems acquiring spatial vocabulary (Jarrold, Phillips, Baddeley, Grant, & Karmiloff-Smith, 2001). The challenge for the conservative hypothesis, however, is to explain why individuals with WS should show errors in, for example, morphosyntax that are not found in typically developing children and why they should show predominantly successful language acquisition when individuals with other genetic syndromes involving mental retardation do not.

The alternative hypothesis, the semantics–phonology imbalance theory, argues that language development in WS takes place under altered constraints. Several atypical constraints have been proposed. These include the idea that individuals with WS have a particular strength in or sensitivity of auditory short-term memory or a particular weakness in lexical semantics. There might be a lag between the development of phonology and semantics, or a problem integrating the two sources of information. The outcome is a system that relies (or has relied at certain points in its developmental history) more on phonological information than semantic information, with certain consequent behavioral impairments. A complication of the imbalance theory is that most of its components are logically independent and not mutually exclusive. Here we consider five subhypotheses separately.

1. Reliance on phonology. This hypothesis suggests that WS language is marked by a particular strength in auditory or phonological short-term memory, to the extent that Vicari, Carlesimo, Brizzolara, and Pezzini (1996) have labeled language in WS as "hyper-phonological." Evidence for this position includes relatively good performance in phonological short-term memory tasks, sometimes within the normal range (Majerus, Palmisano, van der Linden, Barisnikov, & Poncelet, 2001). Phonological short-term memory has been reported as advantaged relative to other aspects of the WS language system (Mervis, Morris, Bertrand, & Robinson, 1999). Phonological short-term memory is a relative strength found in children as young as 2.5 years, the youngest age tested (Mervis & Bertrand, 1997). In older children and adults, phonological fluency is claimed to be relatively advantaged compared with semantic fluency (Temple, Almazan, & Sherwood, 2002; Volterra et al., 1996). In tasks involving the memorization of words, studies have suggested a reduced influence of semantic information in recall (Vicari, Brizzolara, Carlesimo, Pezzini, & Volterra, 1996; Vicari, Carlesimo, et al., 1996), as if participants' recall performance was based preferentially on phonological encoding. Indeed, overall, Bishop (1999) has argued that WS demonstrates the importance of short-term memory for speech sounds in determining the success of language development.

2. Sensitivity of phonology. Auditory or phonological processing may be a relative strength in WS, but there are also suggestions that the representations underlying this ability are atypical. Majerus et al. (2001) argued that good performance on these short-term memory tasks might rely on lower level acoustic rather than phonological short-term storage. Using an event-related potential paradigm, Neville and colleagues (Neville, Holcomb, & Mills, 1989; Neville, Mills, & Bellugi, 1994) found that individuals with WS had activation responses to auditory stimuli that were less refractory and more excitable than those found in controls, a difference that did not extend to the visual modality. Neville et al. (1994) proposed that "the hypersensitivity of the auditory system in Williams subjects may in part underlie the sparing of and the precocious and hyperfluent nature of [their] language, and the fact that this development occurs following abnormal delays in the acquisition of auditory language" (p. 82). When Laing, Hulme,

Grant, and Karmiloff-Smith (2001) investigated phonological awareness in WS, they found significantly lower levels of rhyme detection and phoneme deletion when compared with reading-age matched controls, consistent with differences in phonological representations.

If phonological representations suffer a lack of robustness, this might predict difficulties in generalizing phonological regularities to novel exemplars. Evidence from past-tense and gender-assignment studies discussed previously supports this prediction. In addition, Grant, Karmiloff-Smith, Berthoud, and Christophe (1996) found that, compared with language- and age-matched controls, individuals with WS exhibited an exaggerated difficulty in repeating nonwords with the phonotactics of a foreign language, as if their phonological representations were overly focused on the phonotactics of their native tongue. Finally, in their reading study, Laing et al. (2001) found a reduced ability in the WS group to generalize their reading abilities to pronouncing nonwords, also consistent with insufficiently robust phonological representations.

3. Lexical–semantic impairment. In contrast to the idea that there is a particular strength in phonology in WS, an alternate view is that there is a particular weakness in semantic processing. Volterra and colleagues (Vicari, Brizzolara, et al., 1996; Volterra et al., 1996, 2001) have noted that grammatical problems in WS are especially evident with those aspects of morphology carrying out a semantic function and that individuals with WS perform better than MA-matched controls only in those areas of language in which semantic aspects are not involved. Indeed, we saw previously that early vocabulary development appears to be characterized by reduced constraints from semantics. Several researchers have suggested particular anomalies in the WS lexical–semantic system. Rossen, Klima, Bellugi, Bihrle, and Jones (1996) proposed that anomalous activation dynamics, specifically impaired inhibitory dynamics mediating context effects within the lexicon, lead to imprecise knowledge of concepts in WS and atypical vocabulary usage. These authors pointed to corroborating evidence from event-related potential measurements, in which individuals with WS exhibited an exaggerated signature of semantic processing (Neville et al., 1994). Neville and colleagues viewed these data as consistent with enhanced connections between related lexical items in the auditory modality. Temple et al. (2002) concluded from data on a receptive vocabulary and a naming task that semantic representations in WS are less well specified and that access to the lexicon is fast but inaccurate (although see Thomas, Dockrell, et al., 2002). Temple et al. speculated that this anomaly may lead to consequent problems in morphosyntax, with retrieval failures of irregular forms allowing overextension of regular patterns. Zukowski (2001) suggested that her own data were consistent with the idea that children with WS were aware that irregular verbs had exceptional past-tense forms (and so did not produce regularized forms) but were initially unable to produce the correct irregular form. In her sample, in several cases, further prompting then elicited the correct form.

4. Semantics lags behind phonology. It is possible that a semantics–phonology imbalance could result from a relative lag across developmental time, whereby phonology is in advance of semantics. The only existing study that addresses this possibility did not provide supporting evidence. Nazzi, Paterson, and Karmiloff-Smith (2003) found a delay in aspects of speech segmentation abilities of infants with WS that was only in line with their MA, arguing against any precocious speech processing abilities.

5. Impairment in integrating semantics and phonology. Another version of the preceding hypothesis is that the delay does not exist in the development of the component abilities but in the integration of semantic and phonological information. Several authors have suggested that such integration deficits exist in WS. Karmiloff-Smith et al. (1998) found that when individuals with WS monitored a sentence for a target word, performance was disrupted by syntactic violations except when those violations involved lexically based information. This led the authors to propose that in WS, there is a deficit in integrating semantic information with phonological information in real-time processing. Indeed Frawley (2002; see Thomas, 2002) has argued that WS should be seen primarily as a disorder involving integration deficits between processing modules. Böhning, Campbell, and Karmiloff-Smith (2002) specifically investigated the integration deficit hypothesis using the McGurk effect, in which audio and visual information are combined to determine the recognition of speech sounds. Their results suggested integration was taking place, although more weakly than the controls, and that the WS group relied more heavily on audio than visual information.

In summary, these five subhypotheses argue that language in WS is anomalous (and special) because of a differential balance in the constraints that shape the language development process. However, as these possibilities demonstrate, a precise characterization of such constraints is still a matter of debate. Note that it is at least possible that several of the above hypotheses could conjointly turn out to be true. For example, WS might constitute a case in which there are differences in phonology and in semantics, in a system exhibiting general delay and overlying effects of mental retardation.

Our next goal is to introduce a connectionist model that captures the development of past-tense formation in the normal population and then demonstrate how we may implement each of the above hypotheses by changing the initial constraints of this connectionist system. We will then explore the ability of each manipulation to generate the behavioral patterns previously identified in the WS data.

Connectionist Models of Past-Tense Formation

Connectionist theories of past-tense formation have converged on construing this domain in terms of an associative system that learns the relationship between the phonological form of verb stems and their respective past-tense forms, in the presence of lexical (or perhaps semantic) information. Past tense fits within a wider morphological framework of an inflectional system that learns all such meaning-based modulations of word form, including those for verbs, nouns, and adjectives. However, this full framework has yet to be implemented (see Plunkett & Juola, 1999, for a model combining verb and noun inflection). In the full system, lexical–semantic information would be complemented by a specification of grammatical class and required inflection type at input, to uniquely specify each inflected form.

The main strength of the connectionist approach is that the flexibility of the learning systems in which it is rooted makes it applicable to all languages. The connectionist approach has the potential to show how the apparent distinction between regularity

and irregularity (a distinction without solid cross-linguistic generality) can emerge as a product of learning from patterns of frequency of usage and phonological similarity within a given language (Bates, 1991; MacWhinney & Bates, 1989). Moreover, the models are fully implemented and well-specified computational simulations, making them highly testable and capable of making predictions even at the item level (Hahn & Nakisa, 2000). Current models of past tense still exhibit a number of weaknesses, however. Models have tended to take a piecemeal approach to explaining individual phenomena; novel generalization does not always match human levels; scaled-up models do not reach ceiling performance on acquiring irregular verbs; and simulations have yet to be extended to multisyllabic verb processing.

As we indicated earlier, the connectionist theory is in strong competition with an alternative account of past-tense formation. Pinker's dual mechanism (DM) model (see Pinker, 1991, 1994, 1999; Pinker & Prince, 1988) proposes that regular past-tense forms and irregular past-tense forms are processed by qualitatively different, domain-specific computational mechanisms. In contrast to the purely associative learning mechanisms of the connectionist approach, the DM theory postulates a symbolic, rule-based system dedicated to regular inflections for existing and novel forms and a separate associative memory to store the past-tense forms for individual verbs, including irregular forms. The symbolic rule system operates as a default when a memorized past-tense form is not retrieved from memory. According to the "rule-epiphany theory" (Pinker, 1999, p. 194), during development the rule system is not engaged until the presence of an inflectional rule is identified in the input (presumably by some separate system). The developmental phase of overregularization is then taken to indicate temporary difficulties in coordinating the symbolic and associative mechanisms.

The DM model offers the benefit of explaining the high level of productivity that people show in extending inflectional regularities to novel forms as well as the linguistic parsimony of its account across different forms of inflection (Marcus et al., 1995; although see Hahn & Nakisa, 2000). The DM theory remains computationally unimplemented in specific subdomains such as past tense because its emphasis has been on a wider scale project to characterize all of language processing—simulation of the entire language system is currently unfeasible. The disadvantage of the lack of implementation in any given domain is that it leaves the DM model underspecified, compromising its testability and predictive power. For example, it has yet to be demonstrated that an implemented DM past-tense model could really work in the way it has been described, that its account of acquisition is a viable one, or that the assumptions that make it (descriptively) adequate in accounting for English past-tense performance would not make it insufficiently flexible to account for patterns of inflections in other languages.

The lack of clear specification of a DM model of past-tense formation is particularly problematic when it is applied to developmental disorders. DM explanations of developmental deficits in regular or irregular inflection are usually traced to initial disruption to either the symbolic rule mechanism or the associative memory mechanism. However, such an account leaves open the extent to which, following initial damage to one mechanism, the initially unaffected mechanism would have the ability to compensate across the developmental process. It is a question that simply cannot be addressed with the current level of detail at which the DM model is specified.

The liveliness of the debate between connectionist and DM theories of inflectional morphology should not be underestimated. DM theories have made appeal to a wide range of evidence of differential processing, acquisition, and breakdown of regular and irregular inflection to substantiate the claim that qualitatively different mechanisms are involved. This includes, as we have seen, claims for double dissociations in both acquired and developmental disorders but also evidence from brain imaging in adults and psycholinguistic evidence of, for example, different frequency effects, semantic effects, and priming effects in regular and irregular verbs. However, both theories respect the distinction between phonology-based regularities and word-specific information, and currently, it appears that such a distinction is sufficient to generate disparities in performance between regular and irregular inflection without appealing to qualitatively different types of underlying computation (Lavric, Pizzagalli, Forstmeier, & Rippon, 2001; see Thomas & Karmiloff-Smith, 2002a, for a discussion of simulating past-tense dissociations with connectionist models).

Our approach in this article is to consider how initial parameter changes to an implemented computational model can account for atypical patterns of development. Given that no existing computational model offers a true implementation of the DM theory (Pinker, personal communication, December 2, 2001), our simulations are limited to the connectionist approach. Nevertheless, some implemented models have been viewed by DM theorists as being more consistent with what is intended in DM theory. Thus, Marcus (2001) described Hare, Elman, and Daugherty's (1995) simulation as "effectively implementing" (p. 80) the rule-and-memory model and Westermann and Goebel's (1995) simulation as "the model that comes closest" (p. 72) to an implementation (see also Goebel & Indefrey, 2000; Westermann, 1998, for a related constructivist model). Moreover, Taatgen and Anderson (2002; Taatgen, 2001) have recently put forward an implemented model that is based on a hybrid rule and memory architecture (ACT-R), which has similarities to the DM approach. When we consider the generality of the findings from the connectionist model, these other models give us a more concrete basis to evaluate whether start-state manipulations generalize across models with different architectures or to the DM model itself. Until then, however, our focus is on the connectionist approach, and in the next section, we specify the model that constitutes our baseline of normal development in past-tense formation.

The Normal (Baseline) Past-Tense Model

For our baseline model, we used a version of Plunkett and Marchman's (1991, 1993, 1996) simulation, modified by the addition of lexical–semantic representations in the input layer. In contrast to subsequent, larger models trained on realistic corpuses, Plunkett and Marchman's model used an artificial language representative of the past tense domain. Their model has been criticized on a number of grounds. It is therefore important to understand why we used this model. There are two reasons.

The first reason is that, despite the emergence of subsequent models, Plunkett and Marchman's (1991, 1993, 1996) model remains the one most carefully applied to and rigorously tested against real developmental data. Its strengths and weaknesses are

therefore well known and its behavior thoroughly understood. As a baseline developmental model, it offers a solid (although not perfect) foundation. More important, we do not believe that its simplifications invalidate the conclusions we can draw on the relative effects of atypical developmental constraints.

Plunkett and Marchman's (1991, 1993, 1996) model used a multilayer perceptron to learn to map verb stems to past-tense forms analogous to the mappings found in the English past-tense system. The authors systematically explored competition effects between different inflectional paradigms and the effects of type and token frequency on regular and irregular forms. They demonstrated that in their network, U-shaped learning was, like in real language acquisition, a microphenomenon, occurring not globally but for different irregular verbs at different times and that overregularization can occur without the externally imposed discontinuities in the training set used by Rumelhart and McClelland (1986). A gradual and incremental increase in the size of the training set, equivalent to vocabulary growth in the child, caused the network to undergo a reorganization that resulted in a shift from a mode of rote learning to a systematic treatment of verbs. The reorganization was responsible for the onset of overregularization in the network, as well as generalization of the regular form, and was driven by a critical mass of regular verbs in the vocabulary set (see Marchman & Bates, 1994, for similar evidence for children's regularization). Overall, the authors argued that the model produced a developmental trajectory and pattern of errors comparable with children acquiring the English past tense.

In a detailed analysis, Marcus (1995) criticized the model on several grounds. Perhaps the most serious criticisms were that the model did not show overregularization errors unless half the input vocabulary was regular, conflicting with empirical data of overregularization errors in a child at a point when regulars formed less than 36% of the tokens of the parental input to the child (Marcus et al., 1992), and that the onset of overregularization was still the result of an externally applied discontinuity in the incremental training regime. Plunkett and Marchman (1996) responded that with regard to the onset of overregularization, the relevant proportion of regular verbs is not a function of parental input but of the vocabulary that the child knows—for the child in question, the latter proportion of regular verbs types was substantially higher than that for irregulars, supporting a regular critical mass explanation for overregularization errors. Moreover, Plunkett and Marchman demonstrated that in their model, discontinuities in the training regime were neither a necessary nor a sufficient condition for overregularization errors; such errors bore a nonlinear relationship to the mass of regular verbs in the training set. Overall, Plunkett and Marchman concluded that the assumptions of the model were reasonable and valid and that the model offered important insights into the factors determining language acquisition.

The second reason why we used this model was that its limited training set of 500 triphonemic strings provided a tractable framework within which to examine issues relevant to the role of initial network constraints in simulating impairments in developmental disorders. Modeling involves making sufficient simplifications that the model can practically address its theoretical aims, while attempting to avoid compromising the validity of its assumptions. Plunkett and Juola (1999) have demonstrated that Plunkett and Marchman's (1991, 1993, 1996) model scales up to a more realistic corpus, and thus that its assumptions are reasonable. Although we accept the limitations of Plunkett and Marchman's model, we argue that it nevertheless permits a valid consideration of the effect of network constraints on the trajectory of development, on the relative difficulty of acquiring regular and irregular inflectional mappings, and on the generalization abilities of the network.

The issue of tractability is a real one here, because our framework will ultimately require consideration of the model's developmental performance through variations in 10 dimensions, including variations in the learning rate, the phonological code, the semantic code, the integration of the two codes, and in other parameters such as network architecture, hidden unit levels, type of learning algorithm, and level of internal noise. We believe that the advantage of our approach lies in its systematic and innovative comparison of competing hypotheses concerning the cause of developmental deficits. To a degree, this advantage is gained at the expense of simplifications to the baseline model.

One example of the issue of tractability arises when we come to add lexical–semantic information to the input of the model. The comparison of phonological and lexical–semantic influences is a key focus of our investigation, and a number of past-tense models have incorporated both sources of information in acquiring inflectional morphology (e.g., Cottrell & Plunkett, 1994; Hoeffner, 1992; Joanisse & Seidenberg, 1999; MacWhinney & Leinbach, 1991; Plunkett & Juola, 1999). However, there is no theoretical consensus on how word-specific information should be construed as constraining inflection. Therefore, there is no consensus on how lexical–semantic information should be represented in these models. Some researchers have merely provided lexical information to the network, in which a separate input unit indexes the identity of each word in addition to its phonological form (e.g., Joanisse & Seidenberg, 1999). Others have used distributed representations in which meanings are represented by random binary patterns (e.g., Hoeffner, 1992; Plunkett & Juola, 1999) or constructed according to a predefined semantic feature set (e.g., MacWhinney & Leinbach, 1991). Yet others have proposed structured semantic representations, with exemplars probabilistically clustered around prototypes, again in a distributed format (e.g., Cottrell & Plunkett, 1994; see also Plaut, 1995b; Plunkett, Sinha, Møller, & Strandsby, 1992). However, no systematic comparison exists that charts the effects of each form of representation on the acquisition of inflectional morphology, despite the theoretical issues involved.[2]

Because our investigations involved formulating precise implementations of disruptions to lexical–semantics, we had to commit to a particular form of representation, but the effect of a given disruption could crucially depend on the chosen form. The appropriate response was to run our simulations in triplicate, using three different representational formats for lexical–semantics (localist, arbitrary distributed, prototype distributed) and compare the effect of disruptions in each case. Unfortunately, even this extension of the model could not take into account that the structure of lexical–semantics was likely to develop across the age band from which

[2] For example, similar meanings are sometimes associated with different past-tense forms: *slap–slapped, strike–struck,* and *hit–hit.* Different meanings are sometimes associated with related irregular past-tense forms: *stand–stood* and *understand–understood.* See Pinker (1999) and Ramscar (2002) for discussion.

our empirical data were drawn. Such a consideration, was, however, beyond the scope of our model.

One of the simplifications we made to accommodate these additional simulations was to train the model on the entire corpus rather than on an incrementally increasing vocabulary set. Although an incremental training set has a higher validity, network performance can be sensitive to the composition of the initial small vocabulary set. In consequence, extra replications of the model are required to factor out this source of variability, adding to the simulation time. Plunkett and Juola (1999) have demonstrated that compared with incremental training, whole-corpus training (used by most past-tense models) does not affect the final generalization ability of the network and still affords a valid consideration of the relative difficulty of acquiring regular and irregular inflectional forms. In consequence, the model did not simulate an early period of error-free performance on a small set of regular and irregular verbs. In any case, the data we seek to model is well beyond this period (from 10 years on in the WS sample, 5.5 years on in the typically developing sample), and no empirical evidence exists that would allow us to link errors in inflectional morphology in WS to early vocabulary and, so, constrain the model in an incremental training regime appropriate for WS.

The Target Empirical Data for WS

Figure 1 shows the empirical results from Thomas et al. (2001) for regular and irregular verbs, and generalization to nonce items. Figure 1a compares the performance of individuals with WS against a sample of typically developing individuals on a past-tense elicitation task, plotted against increasing chronological age (CA). The task involved regular verbs (*look–looked*), irregular verbs forming their past tense by a central vowel change (*swim–*

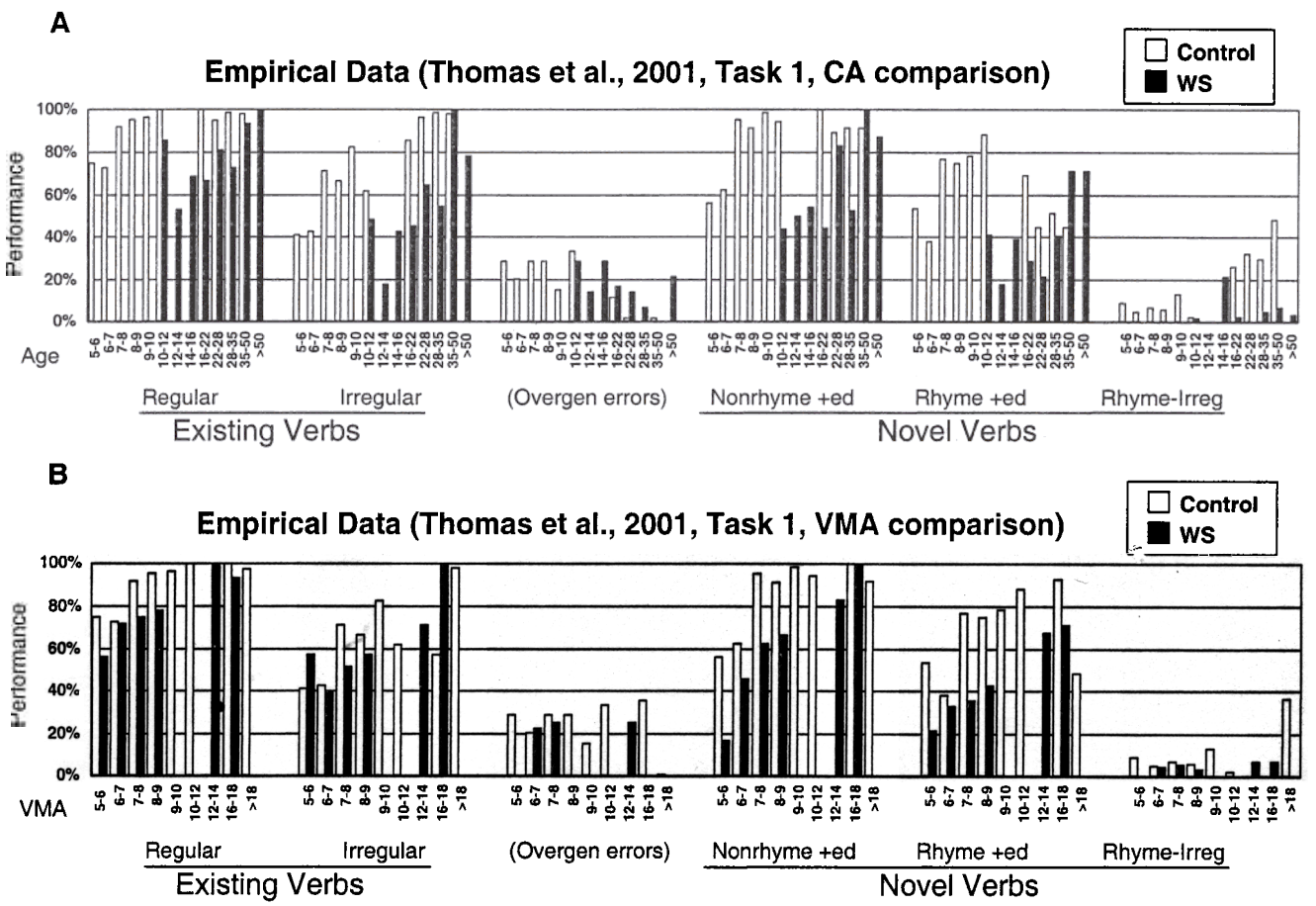

Figure 1. A: Past-tense elicitation performance for individuals with Williams syndrome (WS) and typically developing control participants plotted against increasing chronological age (CA) from Thomas et al. (2001). Data show performance on regular and irregular verbs and three types of novel items: Overgen = overgeneralization; Nonrhyme +ed = regular generalization of novel items not rhyming with existing irregulars (e.g., *brop–bropped*); Rhyme +ed = regular generalization of novel items rhyming with existing irregulars (*crive–crived*; cf. *drive*); and Rhyme-Irreg = irregularization of novel rhymes (*crive–crove*). B: The same data plotted against increasing verbal mental age (VMA). Participant numbers varied across age bins (see Thomas et al., 2001, for further details). The Nonrhyme +ed and Rhyme +ed columns in A and B are adapted from "Past Tense Formation in Williams Syndrome," by M. S. C. Thomas et al., 2001, *Language and Cognitive Processes, 16,* pp. 162–163. Copyright 2001 by Psychology Press, Hove. Adapted with permission.

swam) or a central vowel change and final consonant change (*think–thought*), nonsense verbs not rhyming with any existing irregular verbs (*brop*), and nonsense verbs that rhyme with an existing irregular (*crive*; cf. *drive–drove*). Figure 1b shows the performance of the two groups against increasing verbal mental age (VMA), assessed using a standardized vocabulary test. For CA, regression analyses revealed overall poorer performance in the WS group and a greater deficit for irregular verbs than regular verbs. When vocabulary-based VMA was controlled for, a slight overall disadvantage remained for the WS group, but there was no differential deficit for regulars and irregulars. However, there was a significant interaction whereby generalization showed a greater deficit between control and WS groups than did performance on existing verbs. Both groups generalized the add *–ed* regularity more strongly to nonrhyming nonsense terms than irregular-rhyming nonsense verbs. Figure 1 shows illustrative data with participants split into age bins. Our simulations focus on attempting to capture the differential developmental trajectories (across items and across groups) extracted from these data using regression analyses.

In summary, our main focus is to examine which initial manipulations produce the following pattern: (a) equal delay for regular and irregular inflection and (b) a reduction in generalization to nonce terms. However, given the alternate claims made for WS, we also examine which manipulations produce an exaggerated delay for irregular past-tense formation. The hypotheses that we derived from the WS empirical literature were implemented in the following ways, summarized in Figure 2.

Implementing the Hypotheses for Differential Constraints in WS Language Acquisition

> *Hypothesis 1*: WS past-tense performance is explained by a general delay in language development in individuals with mental retardation.

Under this hypothesis, poorer performance in forming the past tense of existing verbs is explained by general language delay. Reduced generalization is explained as a consequence of slower development or by an untested proposal that individuals with low IQ have a particular difficulty dealing with nonsense terms in inflectional elicitation paradigms. Newfield and Schlanger (1968) reported such a difficulty with a mixed group of individuals with unspecified mental retardation, but close inspection of these data suggests generalization in line with poorer performance on existing verbs. There are two disadvantages with the delay/difficulty-with-nonce-terms approach. First, an appeal to general language delay in WS is unsatisfactory because the delay is not uniform across all aspects of WS language (e.g., grammar is more delayed than vocabulary). Second, the proposal that nonce terms are inappropriate stimuli for this group would deny us the ability to evaluate the productivity of morphosyntax in developmental disorders with mental retardation. However, in the case of WS at least, individuals have not demonstrated problems dealing with nonce terms per se. Karmiloff-Smith et al. (1997) reported that individuals with WS repeated nonce terms more accurately than MA controls (Karmiloff-Smith et al., 1997), and with respect to morphology, Levy and Hermon (in press) found that individuals with WS demonstrated no problems in understanding the task of inflecting nonce terms in a study in which generalization was nevertheless reduced. The hypothesis of general delay was implemented by reducing the overall learning rate in the past-tense model.

> *Hypothesis 2*: WS past-tense formation is based purely on the extraction of phonological regularities and does not use constraints from lexical–semantics.

Given evidence that individuals with WS may be hyperphonological, this hypothesis proposes that deficits occur because of a failure to use lexical–semantic information about specific verbs in learning phonologically based regularities. On the face of it, this hypothesis is contradicted by evidence, because Thomas et al. (2001) reported a stronger semantic effect in irregular past-tense formation in WS than typically developing controls. Nevertheless, in our simulation, we sought to examine the effect of focusing on phonology alone by excluding lexical–semantic information from the training set.

> *Hypothesis 3*: WS past-tense formation combines phonological and lexical–semantic information but is impaired because of atypical phonological representations.

This hypothesis proposes that greater auditory sensitivity in WS leads to phonological representations that include too many features of the native language, because of a focus on exemplar learning. Such representations might cause a delay in acquiring the past tense and form an insufficient basis for robust generalization.

The available empirical evidence does not constrain this hypothesis very strongly. There are suggestions that WS performance may rely on phonological representations that are less abstracted from acoustic detail, that the auditory system in WS has high temporal sensitivity, and that in tasks involving nonword repetition and morphosyntax, generalization is reduced. Several connectionist models have established that reduced similarity in the phonological representations can lead to decreased generalization, for example in the simulations of Rumelhart and McClelland (1986) and Seidenberg and McClelland (1989). However, these models used psychologically implausible coding schemes. To evaluate this hypothesis, we explored the implications of reducing the representational similarity between different phonemes while retaining a psychologically plausible coding scheme. We therefore retained a distributed code that was based on articulatory features, either reducing or increasing the number of available features to provide a continuum of similarity on which phonological representations could vary. In effect, this manipulation stretched or compressed phonological similarity space. We assumed that in WS, such an anomaly would be the outcome of disruptions to an earlier developmental process that derives (native) speech-based representations based on lower level auditory information. Reduced-similarity representational codes used longer vectors to discriminate the full phoneme set. Therefore, as a control, we also examined the effect of increasing phonological redundancy, that is, of using longer vectors that retained the same similarity.

> *Hypothesis 4*: WS past-tense formation combines phonological and lexical–semantic information but is impaired because of less activated or atypically structured lexical–semantic representations.

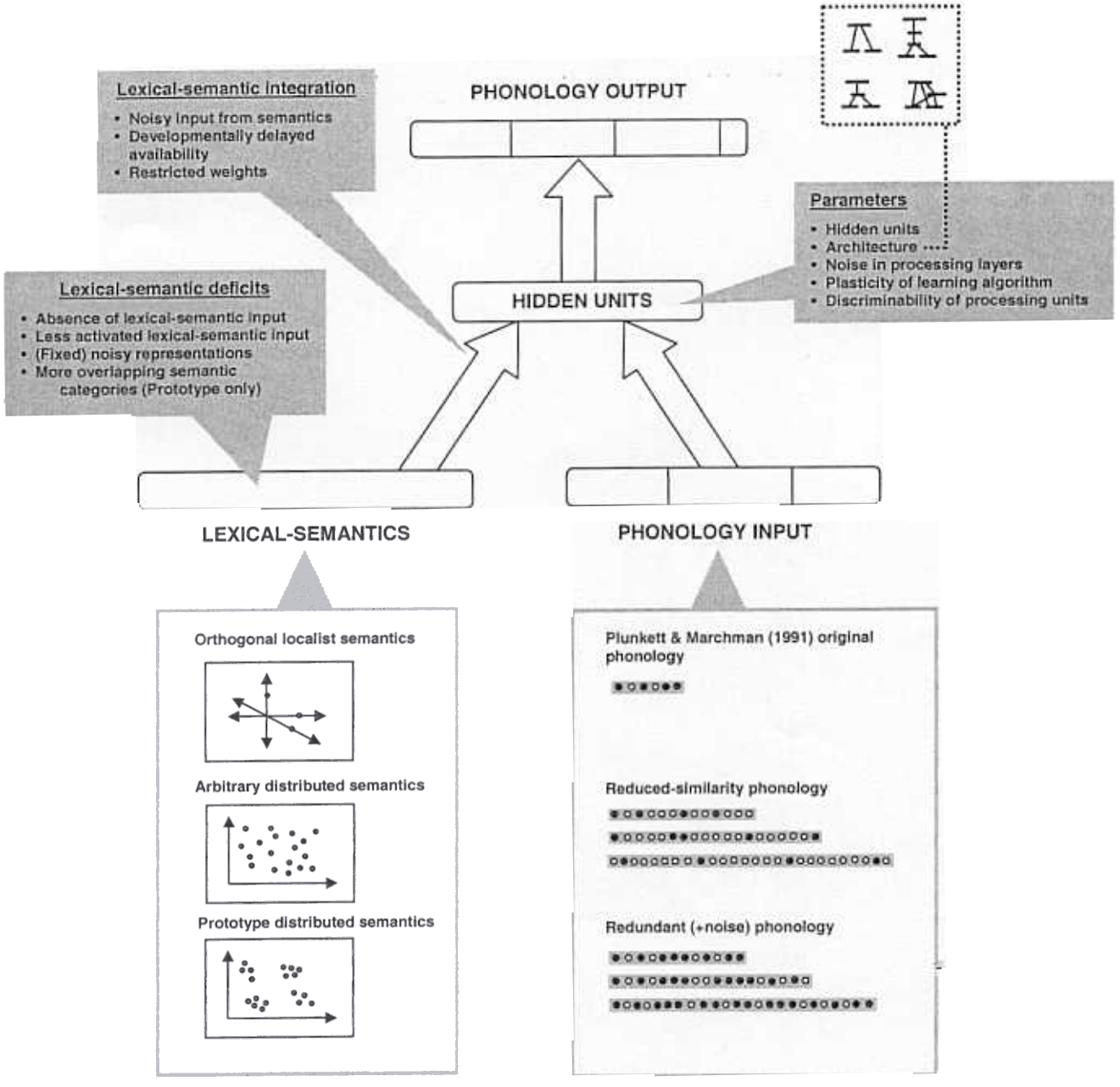

Figure 2. Summary of the model architecture and the main manipulations. These included manipulations to lexical–semantics, phonology, the integration of lexical–semantics and phonology, and a range of background parameters.

Temple et al. (2002) and Clahsen and Almazan (1998) have argued that anomalies in accessing word-specific information may lie at the root of problems in WS morphosyntax, and Pinker (e.g., 1991) has argued by analogy to acquired disorders that lexical–semantic deficits could be implicated in WS past-tense performance. Thomas et al. (2001) maintained that evidence of an increased semantic (imageability) effect in the irregular responses of their WS group might be explained by a reduced lexical–semantic contribution, a claim that we examine in the following simulations. It is not yet clear what the nature of the proposed semantic anomaly in WS might be. It could involve representations that are less strongly activated or more slowly accessed; or it could involve anomalous lexical–semantic organization, such as Temple et al.'s (2002) suggestion of "looser" representations. Clahsen and Almazan (1998, Footnote 18) provisionally argued against a general lexical–semantics deficit, in favor of a more circumscribed deficit in accessing a particular sort of information linked to irregularly inflected words contained in hierarchically structured lexical representations.

We considered three specific implementations of the lexical–semantic hypothesis. First, we reduced the activation levels of the lexical–semantic units by 50%, weakening their contribution to the

past-tense mapping problem. This condition might also be viewed as equivalent to slower on-line access for this information. Second, we generated anomalous organization of lexical–semantic representations by freezing noise into the activation patterns (several noisy versions were created and the results averaged across them). Third, for the version of lexical–semantics that used a prototype organization, we examined the implication of increasing the overlap between the categories, effectively blurring semantic distinctions. In addition, we evaluated Thomas et al.'s (2001) suggestion that increased imageability effects might emerge in past-tense performance because of a reduced contribution from lexical–semantics. Following Plaut and Shallice (1993; see also Plaut, 1995a), we assumed that abstract meanings use fewer diagnostic, context-independent features than concrete meanings. We then evaluated the effect of feature numbers on past-tense performance when the activation of lexical–semantic units was attenuated. Finally, given the links that Pinker (1991, 1994) has drawn between WS and normal adults with acquired lexical–semantic deficits, we contrasted the effect of start-state anomalies with the same damage applied to the end state or adult model, in a direct comparison of the implications of developmental versus acquired deficits (see Joanisse & Seidenberg, 1999).

> *Hypothesis 5*: WS past-tense formation is impaired because of a failure to properly integrate phonological and lexical–semantic information during development.

Under this view, phonology and lexical–semantics may have typical organization, but problems arise in combining these sources of information. Again, we considered several implementations of this hypothesis, given its currently vague specification. First, we introduced on-line noise into the information arriving from the semantic layer, thereby interfering with its ability to reliably constrain the phonological mappings. Second, we reduced the learning rate of the weights connecting lexical–semantic input and the hidden units, again limiting the on-line influence of this information. Finally, as a more transient measure, we delayed the developmental availability of lexical–semantic information such that phonologically based representations were already established prior to the onset of lexical–semantic constraints.

In the introduction, we raised several methodological issues entailed by the strategy of accounting for developmental disorders within the parameter space of a computational model. In the next section, we discuss four of these issues: how one defines a normal model, how individual variability in disorder groups is addressed, the best way to fit simulation results to atypical data, and the flexibility of the chosen model. The reader who wishes to concentrate on the results of the WS simulations may wish to skip this section.

Method

What Defines a Normal Model?

When one constructs a computational model of atypical development, implicitly one must have accepted the validity of a given model of typical development. But how do we decide what the normal model is? A model comprises a range of initial assumptions regarding the network architecture, the nature of the input and output representations, the training regime, the learning algorithm, and so forth. Each of these decisions may be crucial in producing a developmental trajectory that matches the empirical data from the normal population. To what extent are these initial decisions justifiable? Let us take the Plunkett and Marchman (1991, 1993, 1996) model as an example.

The original three-layer architecture was used by Plunkett and Marchman (1991, 1993, 1996) because pilot studies revealed that a network without hidden units could not learn both regulars and irregulars to ceiling. With regard the representational scheme, Plunkett and Marchman (1991) did not aim for an accurate rendition of English phonology but merely for a coding scheme that "reflects a trade-off between accuracy and economy of representation, given the particular set of phonemes used in these simulations" (p. 51). However, Plunkett and Marchman did not suggest why such a trade-off is important. The number of hidden units was determined to optimize performance both on the training set and on generalization. The training regime was chosen to reflect the language available to the child in his or her environment, an issue that Plunkett and Marchman considered quite carefully by varying the nature of the training regime across a range of simulations.

In summary, although some of the model's initial assumptions were justified, several decisions were made to simulate the correct pattern of empirical data. Primarily, Plunkett and Marchman's model is defined as *normal* because it captures the pattern of data for normal children (Plunkett, personal communication, August 16, 1999). This also holds for the vast majority of current models of normal cognitive development. In line with this position, in our simulations, we chose a baseline parameter set to provide an adequate fit to the normal data from Thomas et al.'s (2001) study.

Modeling Individual Variability in Disorder Groups

Clinical populations usually show wide individual differences in the severity of disorder. This is also true of WS (see Thomas et al., 2001, for detailed discussion of relevant data). However, atypical models usually aim to capture group performance. This was our approach in the following simulations. Nonetheless, when individual variability in the behavioral measures of a disorder group is explicitly traced to the underlying disorder itself (rather than to task variability or to normal individual differences), the model should be able to capture this variability within its parameter space. It is not always empirically straightforward to make such distinctions, and the theoretical relationship between individual differences and atypical development has yet to be clarified at a computational level. Elsewhere, we have begun to explore these issues (Thomas & Karmiloff-Smith, 2003), including a consideration of the consequences of this relationship for empirical methodology in studying developmental disorders (Thomas, 2003). Nevertheless, theories of developmental disorders tend to assume in the first instance that each disorder has an underlying cognitive architecture, masked by individual differences and measurement variability. The simulation of group data is motivated by this latter assumption.

Matching the Model to Atypical Data

In the kinds of computational models used to simulate deficits in disorders such as autism, dyslexia, and SLI, there is little independent basis to establish that parameter manipulations are of the correct absolute size. In line with these models, our aim in the following simulations was to take a baseline condition and examine the relative effects of alterations in initial network constraints. We ask the following question: Which constraints push the model's behavior in the direction of the clinical group? However, following from the previous section, the extent to which a given manipulation can vary the severity of the disorder is of interest. We therefore report not just the relative effect but also its outer bound, as far as this could be established. It should be noted, though, that our explorations were predominantly of main effects of individual parameters. As we show, more severe cases may represent higher order interactions between parameters, exaggerating the effect of the each parameter on its own.

Many disorders show overall developmental delays, which presents a problem in matching simulation data to empirical data. For example, for the WS data, should we match past-tense performance against increasing chronological age or verbal mental age (VMA)? Both confound severity of disorder with level of experience, given that the data are cross-sectional. A VMA match enables one to compare the disorder group with controls with an equivalent level of domain-specific achievement, which may thus seem more appropriate. However, the model does not have a VMA, only its performance within the past-tense domain. Indeed, matching on VMA is not theoretically neutral: It places the causal origin of any delay outside the past-tense system. However, in WS, the level of delay varies across different aspects of language, with syntax and even morphosyntax lagging behind vocabulary (Thomas et al., 2001). We decided to match against CA data but also to check that when level of regular performance was controlled for, irregular verbs fell in line with the baseline condition. This comparison is equivalent to Thomas et al.'s demonstration of an equal delay for each verb type. It is a necessary comparison because, as in the empirical data, the lag between regular and irregular past-tense acquisition in the normal population can create an apparent selective deficit for irregular verbs merely under conditions of delay.

How Flexible Is the Model in Capturing Different Patterns of Data?

Many of the initial connectionist models of developmental disorders demonstrated that variation in a single parameter of a model of typical development allowed it to fit data from a target disorder (e.g., reduced hidden units for surface dyslexia in a model of reading [Plaut et al., 1996]; increased hidden units for autism in a model of categorization [Cohen, 1994, 1998]). Do these respective models tell us that behavioral impairments are necessarily caused by alterations in hidden unit numbers, that is, in internal representational resources? Assuming that each normal model is valid, the answer depends on whether the manipulation in question is the *only way* that the model could have fitted the atypical data. However, if many other parameter manipulations also permitted a good fit to the atypical data, then the model would not particularly constrain our theoretical view of the differences that shape the developmental trajectory. Our final aim in the present set of simulations, therefore, was to gauge the background flexibility of the model in fitting the target data. Notice that we never trained the model on the particular pattern of data we wished to fit (see Massaro, 1988). Rather, we altered the initial constraints of the model and examined whether a given atypical pattern of data emerged as a product of training. To explore the background flexibility of the model, we varied parameters including the network architecture, the number of hidden units, the level of noise in activation levels, the plasticity of the learning algorithm, and the discriminability of the processing units.

Simulation Details

Training and Generalization Sets

For our training set, we used the "phone" vocabulary from Plunkett and Marchman (1991, p. 70). The set comprised 500 triphonemic verb stems created by combining consonants and vowels (from a set of 32) into three possible templates conforming to the phonotactics of English (CVC, CCV, VCC). Past tenses could be regular (addition of /d/, /t/, or /^d/ conditioned by the final phoneme of the stem) or irregular. Irregular past tenses were of three possible types: arbitrary, no change, or vowel change. The original set comprised 410 regulars, 2 arbitrary, 20 no-change, and 68 vowel-change, to which we added 8 additional arbitrary past tenses to allow finer graduations of performance (2 verbs only allowing levels of 0%, 50%, or 100%). Following Plunkett and Marchman (1991), the verbs were given a frequency structure. For computational convenience, frequency was implemented by mediating the weight change generated by the difference between the actual output and the target output (Plaut et al., 1996). Pilot simulations showed this to have no marked effect on network performance. The weight change of high-frequency arbitrary verbs was multiplied by 0.9 during a given training presentation and that of low-frequency arbitrary verbs by 0.6. The weight change of all other high-frequency verbs (regulars, vowel change, and no change) was multiplied by 0.3 and of all other low-frequency verbs by 0.1. Although this frequency structure was included in the training set in line with Plunkett and Marchman's artificial vocabulary, we did not consider frequency effects in the results of the simulations.

A large novel verb set of 572 items was created to test generalization. Novel verbs could be of three types. They could either share two phonemes in identical positions with an existing regular verb (410 items), share two phonemes with an existing irregular verb (10 with arbitrary, 76 with vowel change, 20 with no-change verbs), or share only one phoneme with any verb in the training set (56 items). Items sharing two phonemes with no-change verbs were constrained to end in an aveolar consonant (/t/ or /d/). In mapping to the Thomas et al. (2001) empirical data, novel items sharing two phonemes with regular verbs were taken as equivalent to nonrhymes and those sharing two phonemes with irregular verbs as equivalent to rhymes, because in that study, rhyming was defined with reference to irregular verbs.

Lexical–Semantic Representations

In the absence of a consensus in the literature on the appropriate form of lexical–semantic representations (at least, as they interacted with the inflectional morphology system), three codes were created and compared.

Localist. The lexical–semantic input layer contained one unit for each of the 508 verbs in the training set. This scheme embodied the proposal that lexical–semantics constrain inflectional morphology by indexing the identity of words but not representing any similarity between them.

Arbitrary distributed. This and the following scheme embodied the proposal that semantic structure can influence morphological operations. Binary patterns were created at random across 200 arbitrary semantic features. We followed Plaut et al. (1996) in assuming that semantic representations are sparse, such that each word activates few of the possible semantic features, and each semantic feature participates in the meanings of few words. Each feature had a .08 probability of being active in a given meaning. The number of features active in the 508 patterns ranged from 7 to 27. To examine the effect of imageability, we defined *concrete meanings* as those with more than 21 features active, and *abstract meanings* as those with less than 13 features active. This yielded 37 concrete and 51 abstract regular verbs and 9 concrete and 7 abstract irregular (vowel-change) verbs.

Prototype distributed. Twenty prototype patterns were created at random over 200 semantic features, in which each feature had a .15 probability of being active. From 12 of these prototypes, 25 exemplars were generated by flipping each unit from 0 to 1 or 1 to 0 with a probability of .05, and from the remaining 8, 26 exemplars were generated, yielding a total of 508 exemplars. These were assigned randomly to the verbs in the training set.

Note that when novel patterns were applied to the network, no input was supplied to lexical–semantics. It is possible that in human participants, nonsense terms actually create subthreshold *pseudo-meanings*, that is, lexical–semantic activation that aids generalization. In pilot simulations, lexical–semantic representations on their own were unable to provide a robust basis for generalization of past-tense regularities (add *–ed* generalization levels: 13% for localist codes, 5% for arbitrary, and 1% for prototype). For simplicity, generalization was examined purely on the basis of phonological similarity. Pilot simulations also revealed that when past-tense forms were computed solely on the basis of the lexical–semantic similarity, there was no difference in performance between those designated as regular or irregular past tenses, suggesting that the encoding of meaning was neutral to this distinction.

Training and Testing Regime

Networks were initialized with connection weights randomized between ± 0.5. Networks were trained by exposure to the entire training corpus for 5,000 epochs with a learning rate of 0.01 and a momentum of 0. Pattern presentation was in random order without replacement. Network performance on training and generalization sets was tested at 10, 25, 50, 100, 250, 500, 1,000, 2,000, and 5,000 epochs. Weight changes were calculated using the backpropagation algorithm (Rumelhart, Hinton, & Williams, 1986) and the cross-entropy error measure (see Hinton, 1989).

Testing was performed using a nearest neighbor procedure. For a given output, the pattern of activation in each phoneme position was converted into the pattern for the closest existing phoneme, using Euclidean distance. This "cleaned-up" version was then tested to see whether it was identical to the target output or corresponded to several possible alternatives (regularization, irregularization, no change, blend, or other). Reported performance scores are the proportion of outputs whose cleaned-up versions corresponded to the target output. Under a more stringent test, the cleaned-up version was only accepted if the summed squared error between the target and original output was less than a threshold, set at 0.1. Although lowering scores during learning, this more stringent test did not change the qualitative pattern of the data, and so the results are not reported here.

Parameter Manipulations

Delayed development. Delayed development was simulated with a slower learning rate of 0.001.

Pure phonology. The pure phonology hypothesis was examined by training the network without the lexical–semantic component of the model. Mappings were learned between phonological representations of the verb stem and past-tense form alone.

Phonological representations. Plunkett and Marchman's (1991, see Table 1, p. 31) original coding scheme used triphonemic strings, with each of the three positions represented by a distributed coding of the phoneme over 6 units. One unit represented consonant or vowel, 1 unit represented voicing, 2 units represented manner of articulation, and 2 units represented place of articulation. The output layer included 2 additional units to represent four possible past-tense inflections, /d/, /t/, /^d/, and no inflection, coded in a distributed fashion. The phonological input layer thus comprised 18 units and the output layer comprised 20 units. We refer to this as a six-bit representational scheme.

Phonology with reduced similarity. To contrast with the 6-bit representational scheme, a second distributed code was designed, on the basis of standard linguistic categorizations (Fromkin & Rodman, 1988), and coded over 19 binary features.[3] Because this scheme used a longer vector to represent the 32 phonemes, the similarity between each phoneme was less. Two additional distributed coding schemes were created. One used 13 features and offered an intermediate level of similarity. One used 26 features and offered a further reduction in similarity. The 13-bit code was created by deleting 6 of the features of the 19-bit code (+coronal, back, nasal, lateral, central, diphthong). The 26-bit code was created by adding 7 dummy features, 3 producing extra contrasts between vowels and 4 producing extra contrasts between consonants. Similarity was reduced because few of the phonemes possessed these additional features. We make no linguistic claims concerning the form of these representations. They merely create a similarity gradient that allowed us to investigate the impact of reduced similarity on past-tense performance, while retaining the principle of distributed feature-based representations of phonology. Accordingly, the mean angle between two phonemes increased from 55.4° (SD = 16.3°) in Plunkett and Marchman's original code to 57.3° (SD = 16.2°) in the 13-bit code, 61.4° (SD = 14.4°) in the 19-bit code, and 64.7° (SD = 13.4°) in the 26-bit code, where 90° signifies zero similarity. Finally, extended representations of the inflectional affix were assigned to the new codes, distributed over 4, 5, and 6 bits for the 13, 19, and 26-bit codes, respectively.

Phonology with increased redundancy. The above reductions in similarity caused the phonological input and output representations to take on increasing size (input/output units: 6-bit, 18/20; 13-bit, 39/42; 19-bit, 57/62; and 26-bit, 78/84). To check what effects an increase in the size of the input and output layers would have in the absence of marked changes in similarity, we constructed three additional coding schemes that possessed increasing levels of redundancy. Accordingly, Plunkett and Marchman's (1991) original code for each phoneme was duplicated n times. Duplication of the codes also introduced some noise, whereby in the duplication process, a bit was flipped from 1 to 0 or 0 to 1 with a probability of .2. Noise was added under the assumption that realistic representations with higher redundancy are unlikely to be pure replications of the concise format. Once the duplicated code for a given phoneme was created, this code was then used across all simulations. Three values of n were used, 2, 3, and 4, to be referred to respectively as times2, times3, and times4. This created networks with phonological input/output layers of 36/40, 54/60, and 72/80 units, respectively.

Lexical–semantic deficits. Attenuated semantic input was implemented by reducing the activation of units by 50%, so that binary features were either 0 or 0.5. Anomalous semantic organization was implemented by adding noise to the normal lexical–semantic codes, with a Gaussian distribution and a variance of 0.1. Activation levels were cropped so that they could not exceed the maximum and minimum levels of 1 and 0. The noisy version was then frozen and used as usual in training. Three versions were created for each lexical–semantic code, and results were averaged across them. For the prototype code, blurring of semantic categories was implemented by applying an additional probability of .05 that features would flip their state in each exemplar, thus increasing the overlap of the prototype structure.

Integration deficits. An integration deficit was simulated in three different ways. First, Gaussian noise was added to the activation levels arriving at the hidden layer from the lexical–semantic input units. Performance was examined under noise levels with variances of 0.05, 0.1, and 0.2. Second, delayed developmental availability was simulated by deactivating the lexical–semantic layer until a certain point in training, either epoch 50, 250, or 1,000. Third, the learning rate of the connections between the lexical–semantic layer and hidden units was reduced compared with that in the rest of the network. Four rates were initially examined, at 50%, 25%, 10%, and 1% of the learning rate in the rest of the network (0.01).

Background Parameters

Architecture. In addition to the standard three-layer architecture, four other architectures were examined. We used a two-layer network, a four-layer network with 25 hidden units in each layer (splitting our normal level of 50 between two hidden layers), a four-layer network with 50 hidden units in each layer (duplicating our normal level in each layer), and a fully connected architecture in which the input layer was directly connected to the output layer as well as through a single hidden layer with 50 units.

Hidden units. The effect of hidden unit resources was examined in the three-layer net. Values of 10, 20, 50, 200, 500, and 1,000 were used, with 50 units constituting the normal condition. Note that 500 units corresponds roughly to 1 unit per verb in the training set, and 1,000 units to 2 per verb.

Noise in general activation levels. Gaussian noise was added to the activation levels of units in the hidden layer, with variances of 0.025, 0.05, 0.1, and 0.2.

Plasticity of learning algorithm. In the baseline condition, the backpropagation algorithm was used with the cross-entropy (CE) error measure. The algorithm was also run with the summed squared error (SSE) measure. The latter version produces less weight change for a unit when it is

[3] Binary features were as follows: sonorant, consonantal, syllabic, continuant, voiced, labial, anterior, +coronal, back, strident, nasal, lateral, −coronal, high, central, low, rounded, tense, and diphthong. For diphthongs, vowel height was coded from where the first vowel starts (i.e., if the diphthong starts high but moves to low, it was coded as high).

committed to an erroneous response and is therefore a less plastic algorithm. In addition, we used the SSE measure with the sigmoid prime offset (SPO; Fahlman, 1988; see Bullinaria, 1997). This parameter serves to increase the plasticity of the network by adding a fixed constant to the derivative of the sigmoid activation function. Pilot studies suggested that the SPO, set at 0.1, offered a level of plasticity intermediate between CE and standard SSE.

Discriminability of processing units. The discrimination ability of a processing unit corresponds to the steepness of its sigmoid activation function. The output of a processing unit is defined by the equation

$$Output = \frac{1}{1 + e^{-(net\ input/temperature)}},$$

where net input is the summed activation arriving at the unit (including its bias), and where the temperature (T) parameter controls the steepness of this function (see, e.g., Hinton & Sejnowski, 1986). Three values were used: T = 1.00 (normal), T = 0.25 (high discriminability), and T = 4.00 (low discriminability).

Interactions

In addition to the main effects of each of the preceding variables, we also examined two interactions. First, we looked at whether alterations in the lexical–semantic code moderated the effect of changes in phonology. Second, because we had altered the basic structure of Plunkett and Marchman's (1991) model by the addition of lexical–semantics, we checked how this had altered the behavior of the system by comparing the influence of the background parameters in the presence or absence of lexical–semantic information.

Replications

Data were averaged over six runs of each network using different randomized starting weights. Standard errors were typically small, and for clarity, some figures do not include error bars. Differences reported in these figures were all reliable.

Results

The Normal Model

In agreement with the previous discussion, we chose as our normal or baseline model the parameter set that gave a best fit to the normal trajectory of development and to the final levels of performance. When lexical–semantic representations were added to Plunkett and Marchman's (1991) original design, one result was that irregularization of novel irregular-rhyming verbs (*crive–crove*) disappeared from network performance. Therefore, we chose a more redundant phonological scheme as our baseline, which restored this aspect of the normal profile (the times4 code; compare with "P&M original" in Figure 5). The normal empirical data also exhibit a reduction in the generalization of the add *–ed* regularity to novel rhymes at higher ages. To capture this characteristic, we selected the localist semantic code that also produced this decline at higher levels of training (see Figure 7). More important, the baseline condition exhibited the standard superiority effect for regular past tenses over irregulars found in the elicitation performance of older children. Although the network exhibited an initially rising, then declining, level of overregularization errors, this pattern should not be construed as the characteristic U-shaped profile found in younger children. In the network, we included no early phase of good performance on a small set of irregular verbs. For this initial phase, an incremental training regime is required in these architectures. In the whole-corpus model, some irregulars appear initially only in the regularized form. In comparison with the empirical data, the relevant phase of comparison for overregularization errors is the reducing tail (see Figure 13). Finally, in architectural terms, the baseline condition used a three-layer network with 50 hidden units. In all figures, the normal–baseline condition is represented by a solid black line. Arrows attached to the baseline condition indicate the direction of change to fit the WS profile of Thomas et al. (2001). In terms of changes in irregularization, the empirical data demonstrated no significant effect, so no arrow is attached to this line.

Hypothesis 1: Delay

Figure 3 shows the result of reducing the learning rate by a factor of 10, compared with the baseline condition. Data are reported for the network's performance on irregular vowel-change verbs, because no arbitrary or no-change verbs were used in the

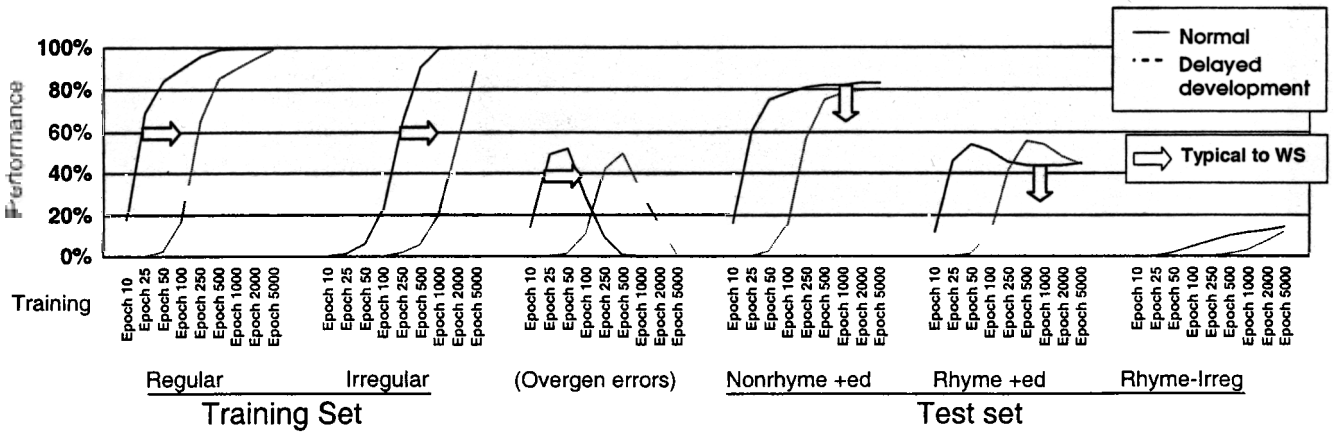

Figure 3. Effect of a slower learning rate in the model, compared against the normal developmental trajectory. Arrows show the direction of change required to fit the Williams syndrome (WS) profile. Overgen = overgeneralization; Nonrhyme +ed = regular generalization of novel items not rhyming with existing irregulars (e.g., *brop–bropped*); Rhyme +ed = regular generalization of novel items rhyming with existing irregulars (*crive–crived*; cf. *drive*); and Rhyme-Irreg = irregularization of novel rhymes (*crive–crove*).

Thomas et al. (2001) or Clahsen and Almazan (1998) elicitation tasks. In addition, data are shown for novel verbs that rhyme with existing irregular verbs (rhymes; e.g. *crive*, *drive*), for which both regularization and irregularization rates are reported (*crived*, *crove*) and for novel verbs that do not rhyme with irregular verbs (non-rhymes; e.g., *brop*), for which only regularization rates are reported (e.g., *bropped*).

Figure 3 demonstrates that although slow learning introduced the requisite delay, there were no implications for generalization. The reduction in generalization found in the WS data cannot be generated by slower acquisition of the domain alone. In terms of the outer bounds of this parameter, at extreme levels, a reduced learning rate prevented successful acquisition of the past-tense domain within the developmental time window.

Hypothesis 2: A Hyperphonological Morphology System

We evaluated the performance of the baseline network against that of a network deprived of lexical–semantic input, which therefore must focus on phonological regularities alone. (This comparison is included as part of Figure 7, labeled as *No lexical–semantics*.) A hyperphonological approach led to accelerated acquisition of regular verbs and delayed acquisition of irregular verbs compared with baseline. In addition, add *–ed* generalization to novel nonrhymes showed a tendency to drop away at higher levels of training in the absence of lexical–semantics. (As an illustration of the variability in the simulation data, the difference between baseline and hyperphonological conditions in generalization of novel nonrhymes at 5,000 epochs was 7.2%; two-tailed t test over six replications: $p = .0003$.) Last, add *–ed* generalization to novel rhymes was reduced, and irregularization of rhymes increased. Such a combination suggests that in the absence of lexical–semantics, irregular verbs play a more salient role in structuring the internal representations of the network. One role of lexical–semantic information in the normal network is thus to partition away knowledge about irregular verbs such that it has a reduced influence on generalization. Compared with the WS data, however, the hyperphonological network failed to exhibit a general delay in acquiring existing past tenses or a general reduction in generalization. In terms of the outer bounds of this parameter, phonology-based learning did not prevent successful acquisition of the training set.

Hypothesis 3: Atypically Structured Phonological Representations

We explored two manipulations to phonology. First, Figure 4 demonstrates four graduations of decreasing similarity between the component phonemes (from the high similarity of the 6-bit representation to the low of the 26-bit representation), against the baseline condition. A reduction in similarity produced three relative effects. First, it delayed the acquisition of regular past tenses. Second, it accelerated the acquisition of irregular verbs. Third, there was a reduction in regular generalization with an increase in irregularization. In short, reduced similarity had a pro-irregular effect. We then explored changing the redundancy. Figure 5 demonstrates a comparison of the baseline condition (most redundant) against conditions of decreasing redundancy. Decreasing redundancy tended to delay both regulars and irregulars but irregulars much more so: Regularization of rhymes increased and their irregularization decreased. In short, reduced redundancy had an anti-irregular effect. The interesting comparison is if we vary both of these parameters at once. Figure 6 depicts the baseline condition against a phonological representation that has both decreased levels of redundancy (i.e., a shorter vector) and decreased levels of similarity, with the lexical–semantic representations held constant. This reveals a delay for regular and irregular verbs that is now similar and an overall general reduction in generalization. This accords with the target pattern of deficits in the WS data. In this combined manipulation, the relatively anti-irregular effect of re-

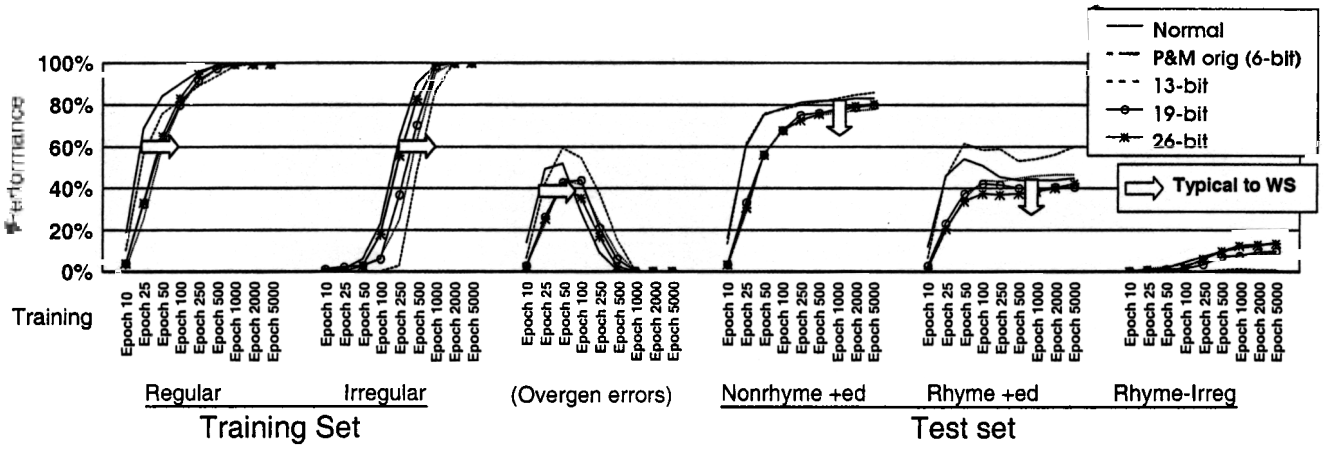

Figure 4. Effect of reducing the level of between-phoneme similarity in the phonological representations. The 6-bit code had the highest similarity, whereas the 26-bit code had the lowest. The normal condition was a redundant version of the 6-bit code. Arrows show the direction of change required to fit the Williams syndrome (WS) profile. P&M orig = Plunkett and Marchman (1991); Overgen = overgeneralization; Nonrhyme +ed = regular generalization of novel items not rhyming with existing irregulars (e.g., *brop–bropped*); Rhyme +ed = regular generalization of novel items rhyming with existing irregulars (*crive–crived*; cf. *drive*); and Rhyme-Irreg = irregularization of novel rhymes (*crive–crove*).

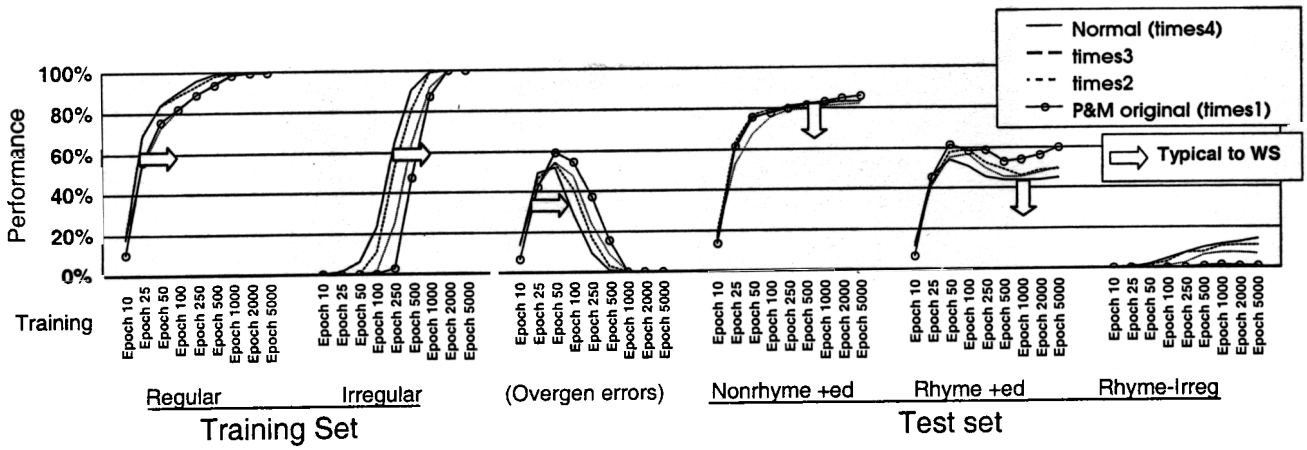

Figure 5. Effect of increasing the level of redundancy in the phonological code. Times4 is the most redundant and times1 is the least redundant. Arrows show the direction of change required to fit the Williams syndrome (WS) profile. P&M = Plunkett and Marchman (1991); Overgen = overgeneralization; Nonrhyme +ed = regular generalization of novel items not rhyming with existing irregulars (e.g., *brop–bropped*); Rhyme +ed = regular generalization of novel items rhyming with existing irregulars (*crive–crived*; cf. *drive*); and Rhyme-Irreg = irregularization of novel rhymes (*crive–crove*).

duced redundancy has overridden the relatively pro-regular effect of reduced similarity, leaving overall delay. The result suggests that atypical phonology could in principle explain WS past-tense performance but only under a particular manipulation. In terms of outer bounds, within the variations considered, neither similarity nor redundancy prevented eventual successful acquisition of the domain.

Hypothesis 4: Lexical–Semantic Anomalies

In this section, we first briefly consider the implication of using different normal formats of representation for lexical–semantics, from localist to distributed random binary patterns to distributed patterns based around a prototype structure. Figure 7 shows a comparison of these three conditions, along with a network trained in the absence of lexical–semantic input. The implications of lexical–semantic representations were as follows. The addition of such information delayed the acquisition of regular past tenses but accelerated the acquisition of irregulars. The more systematic the structure within semantics, the greater the effect was. As the semantic input increasingly predicted the output form, regularities between phonological input forms and outputs became less influential in determining relative success. (Recall that in a pilot network in which past-tense form was driven only by semantic input, there was no difference between regulars and irregulars.) In terms

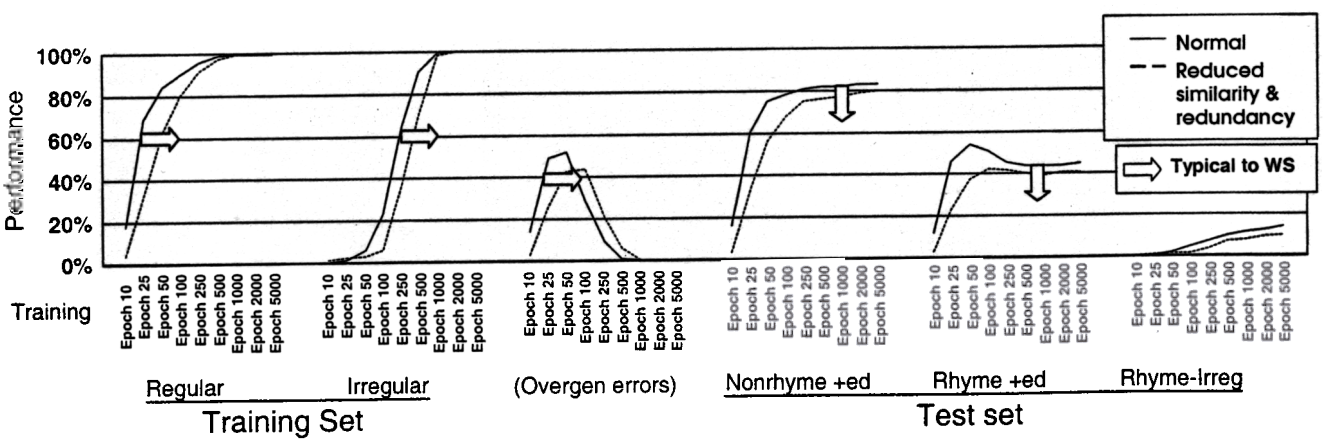

Figure 6. Comparison of the baseline condition (times4) against a phonological representation with reduced similarity and redundancy (19-bit). Arrows show the direction of change required to fit the Williams syndrome (WS) profile. Overgen = overgeneralization; Nonrhyme +ed = regular generalization of novel items not rhyming with existing irregulars (e.g., *brop–bropped*); Rhyme +ed = regular generalization of novel items rhyming with existing irregulars (*crive–crived*; cf. *drive*); and Rhyme-Irreg = irregularization of novel rhymes (*crive–crove*).

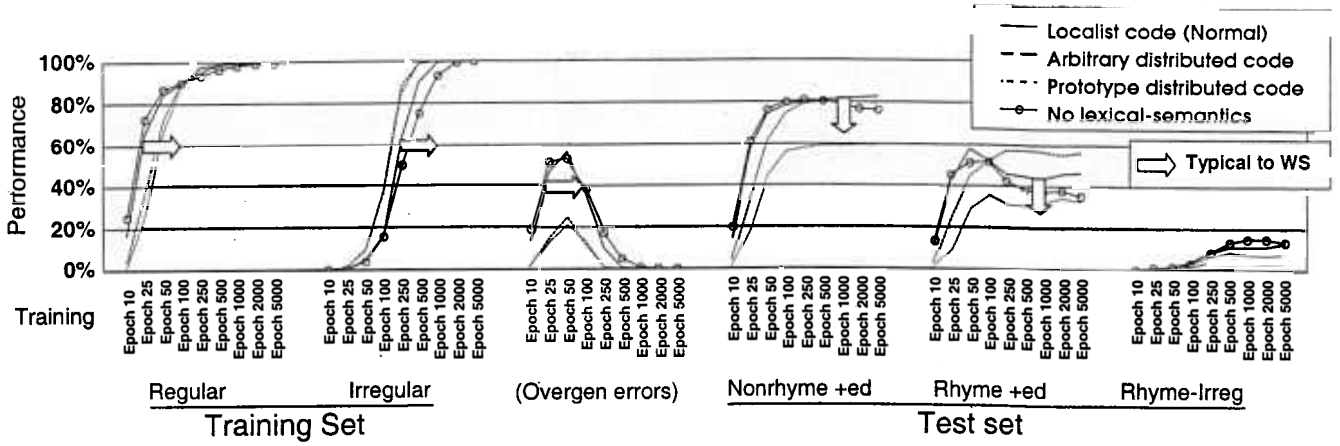

Figure 7. Effects on past-tense performance of changing the encoding of the lexical–semantic representations and of excluding lexical–semantics from the network. Arrows show the direction of change required to fit the Williams syndrome (WS) profile. Overgen = overgeneralization; Nonrhyme +ed = regular generalization of novel items not rhyming with existing irregulars (e.g., *brop–bropped*); Rhyme +ed = regular generalization of novel items rhyming with existing irregulars (*crive–crived*; cf. *drive*); and Rhyme-Irreg = irregularization of novel rhymes (*crive–crove*).

of generalization, regular generalization was improved by the addition of localist and arbitrary distributed vectors identifying each verb. However, as soon as a systematic, prototype similarity structure was included, generalization dropped markedly (see Cottrell & Plunkett, 1994, for a similar result in a recurrent system).

Next, we consider the impact of anomalies to the function/structure of lexical–semantics. In this context, recall that some authors have compared the developmental disorder of WS with the behavior found in some acquired deficits: for example, in fluent aphasia and Alzheimer's disease (e.g., Pinker, 1991, 1994). Here, therefore, we compared developmental lexical–semantic anomalies (applied to the initial conditions of the network) against acquired anomalies (applied to the fully trained network). Figure 8 shows these comparisons, separately for the localist, arbitrary distributed, and prototype distributed coding schemes. Normal performance was compared against (a) a system in which activation levels from lexical–semantics were attenuated by 50%; (b) a system that received input from a lexical–semantics system with anomalous organization, disordered by frozen-in noise; (c) in the case of the prototype representations, a system in which the overlap between semantic categories had been increased, blurring semantic distinctions; and (d) a system that had experienced a full lesion of the lexical–semantic system.

The results revealed that none of the *developmental* deficits prevented ultimate success in learning the mappings for this domain (compare each of the 5,000 epoch columns). In contrast, *acquired* deficits to lexical–semantics demonstrated larger deficits to irregular performance than regular performance (replicating an earlier finding by Joanisse & Seidenberg, 1999, using only the localist code). The effect was more marked for arbitrary distributed and prototype distributed codes in which the system relied more heavily on this information to generate its output. The biggest acquired dissociation was apparent for the arbitrary code, for in this network, the system relied on lexical–semantic information to drive irregular past-tense formation but not regular past-tense formation, whereas for the prototype code, the system relied on it to drive both. Compared with these acquired deficits, developmental anomalies tended to lead to *delays* in acquisition. In this model, therefore, although developmental and acquired deficits are clearly related, they do not produce the same effects.

Delays caused by lexical–semantic anomalies were on the whole restricted to irregular verb formation (compare each of the 250 epoch columns), with the exception of noise added to localist representations. Two further points are of note with regard to the WS profile. First, for all lexical–semantic codes, generalization was reduced by anomalous structure caused by frozen-in noise. However, for arbitrary and prototype schemes, there were no implications of this manipulation for the acquisition of existing verbs, and with the localist code, frozen noise was associated with a delay for regulars but accelerated irregular acquisition. Therefore, frozen noise could not account for the overall-delay/reduced-generalization WS pattern. Second, for the prototype lexical–semantics, imposing a greater overlap between semantic categories generally served to *improve* performance, suggesting that the tight semantic structure of the original actually impaired performance. Given the levels of generalization exhibited by normal participants in empirical studies, this appears to set bounds on the extent to which semantic information can influence inflection.

In summary, lexical–semantic anomalies in isolation did not appear to allow the system to capture the overall-delay/reduced generalization profile of the WS empirical data. Conversely, such anomalies did appear consistent with the sometimes-reported pattern of differentially delayed irregular performance.

Finally in this section, we consider differential effects of imageability on past-tense performance. Recall that when Thomas et al. (2001) examined the effect of this dimension, they found a differential pattern between the overall WS group and controls. Neither group showed imageability effects for regular past-tense performance. However, the WS group (but not the controls) demonstrated both poorer performance on abstract irregular verbs than

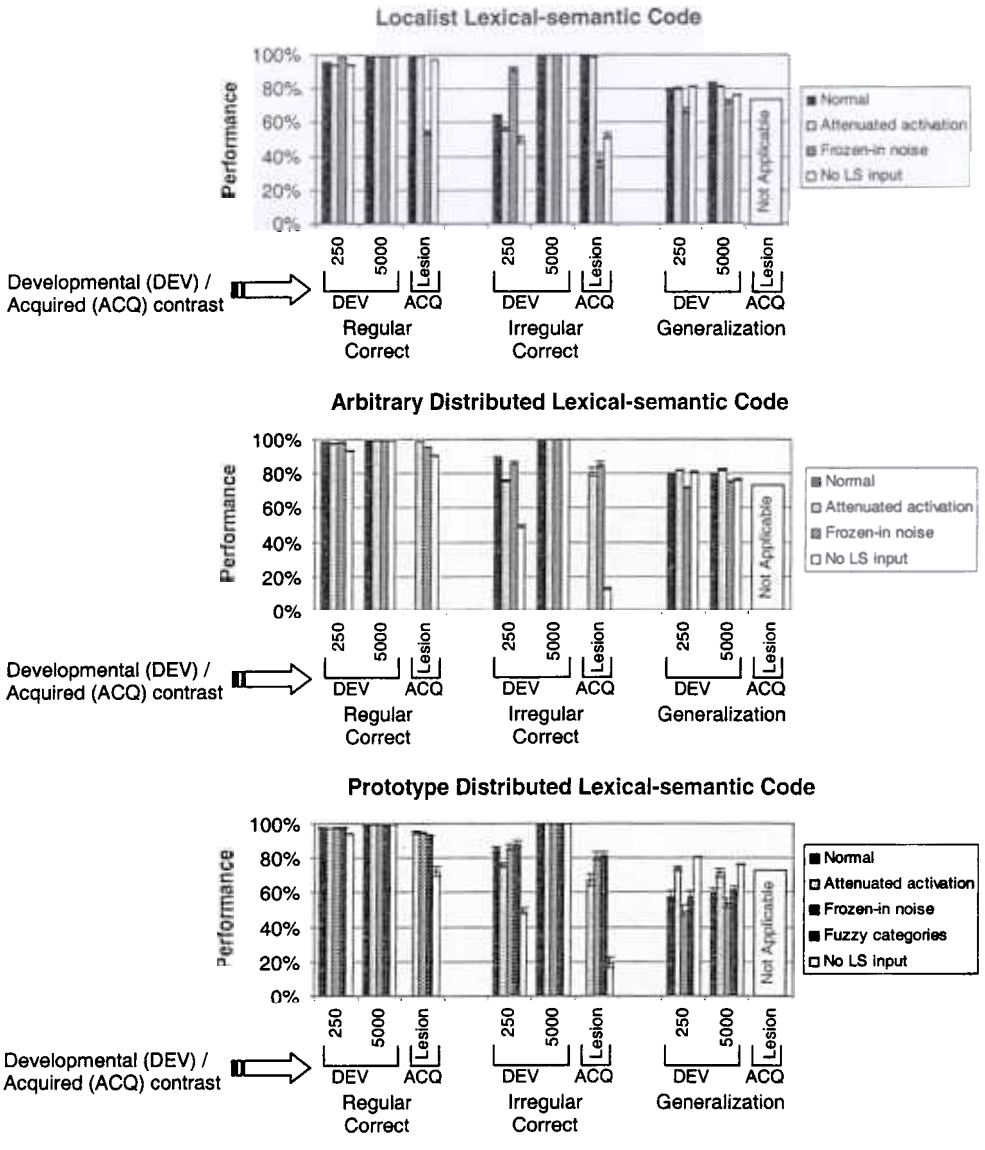

Figure 8. Comparisons of lexical–semantic (LS) deficits applied prior to training (developmental) or as a network lesion at the end of training (acquired). For deficits applied prior to training, performance is shown midway through the training process (250 epochs) and at the end of training (5,000 epochs). Scores shown are for performance on regular verbs, irregular verbs, and add *–ed* generalization of nonrhymes. Results are shown separately for each type of lexical–semantic code. Error bars show standard error across network replications. Note that acquired deficits to lexical–semantics did not affect generalization because novel inputs were purely phonological. Developmental deficits did affect generalization because they constrained the developing structure of the internal representations.

concrete and a raised level of overregularization errors for abstract verbs. We explored whether such an effect might be traced to a *reduced* influence of lexical–semantics, as proposed by Thomas et al.

We used the arbitrary distributed coding scheme, and followed Plaut and Shallice (1993) in assuming abstract verbs to be those with fewer semantic features active. We then compared the performance of the normal network and the network trained with attenuated lexical–semantics, at the midway point in training (250 epochs). The results are shown in Figure 9, along with the empirical data from Thomas et al. (2001). p values are for related-samples t tests, for the network across its six replications.

Although the model's performance was at a higher level than the human data, the simulations replicated the pattern on irregular verbs: presence of an imageability effect for irregular verbs in the atypical but not the typical network and significantly more overregularization errors for abstract irregular verbs than concrete in the atypical network but not the typical. Too much should not be read into the p values for the model. Further replications would cause the imageability effect on correct irregular verb performance

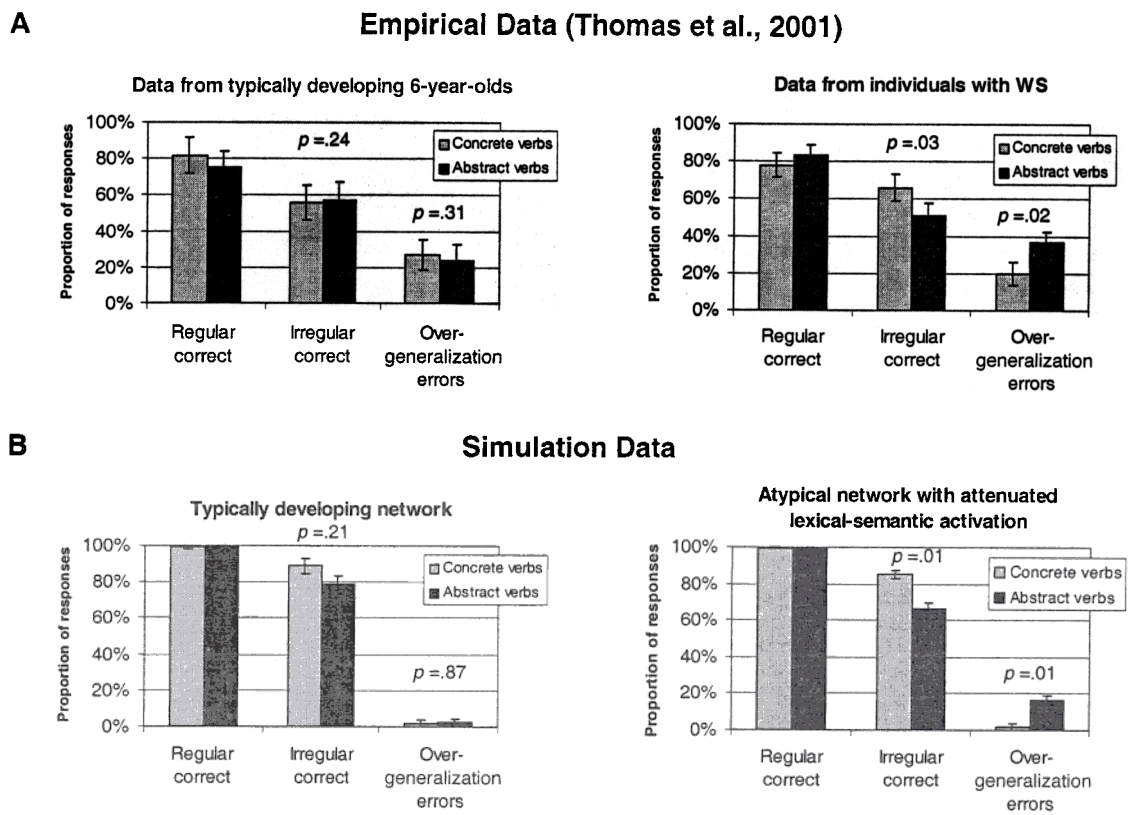

Figure 9. A: Empirical data from Thomas et al. (2001), comparing performance on concrete and abstract verbs in a past-tense elicitation task. Results are shown for performance on regular verbs, irregular verbs, and proportion of overgeneralization errors on irregulars (e.g., *think–thinked*). B: Simulation data comparing the performance of a network with normal and with weakened semantics on concrete and abstract verbs. Error bars represent standard errors of group means for the empirical data and standard errors across the six network replications for the simulation data. WS = Williams syndrome.

in the typical network to reach significance as well. However, additional replications would not, as far as we can tell, lead to a significant imageability effect in overregularization in the typical model. Networks never showed imageability effects in regular verbs (either earlier or later in training) because, as we have seen for the arbitrary distributed code, it is only the irregular verbs that rely on lexical–semantic information, a solution that the network learns for itself during training.

Imageability effects were exaggerated when lexical–semantic activation was attenuated for the following reason. Lexical–semantic information is used to disambiguate irregular verbs from regular verbs in the network and so protect them from overregularization. This ability depends on the net input of lexical–semantic information, in this case number of features active. In the normal network, both concrete and abstract verbs have sufficient net input to allow disambiguation and prevent overregularization. In the attenuated condition, concrete verbs retain sufficient input, but now abstract verbs do not. The result is an increased imageability effect.

In summary, these simulation results support the interpretation that at least one of the features of Thomas et al.'s (2001) WS past-tense data may relate to a reduced lexical–semantic influence on inflection, but that this is insufficient to explain the full pattern.

Hypothesis 5: Integration Deficit

The integration deficit was examined using three manipulations: addition of noise to activations arriving from lexical–semantics, a delay in the developmental availability of lexical–semantic information, and a restraint on the learning rate of the connections from the lexical–semantic layer. Figure 10 compares these conditions. For the latter two, the pattern of results was similar to that achieved by the lexical–semantic deficits considered in the previous section. However, the addition of noise disrupted the network in learning the function relating stems to past-tense forms. There was a uniform delay for regulars and irregulars, as well as an overall reduction in generalization. If the integration deficit is construed in these terms—as disruptive—then this condition also allowed the model to capture the WS data profile. In terms of the outer bounds of this manipulation, additional noise was sufficient to impair successful acquisition.

Interactions Between Phonology and Semantics

In addition to the main effects of the phonological and lexical–semantic manipulations, we also explored whether an interaction existed between them. Specifically, did a manipulation such as

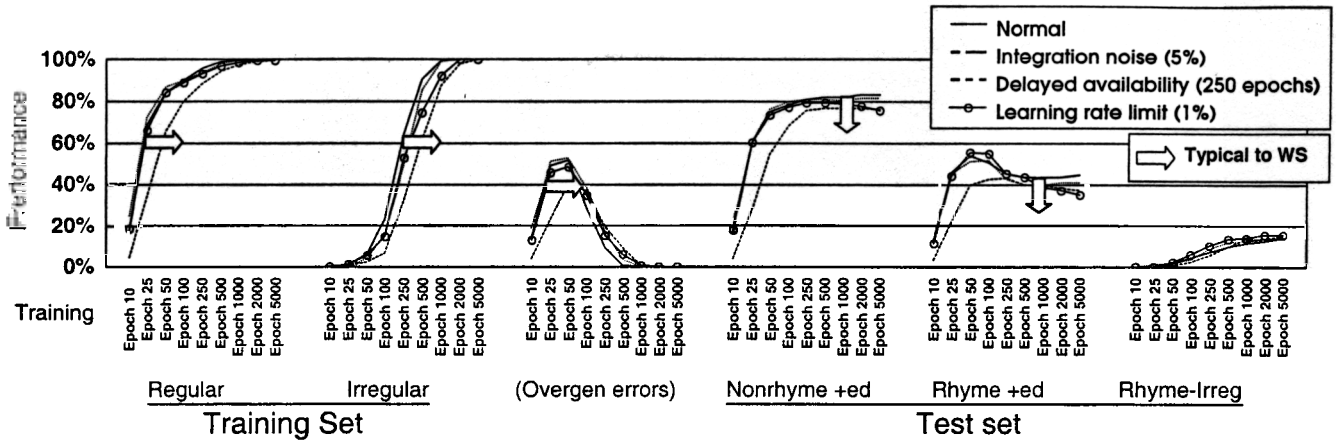

Figure 10. Effect of different types of lexical–semantic integration deficit: adding noise into the signal, developmentally delaying its availability, or restricting the weight change in these connections. Arrows show the direction of change required to fit the Williams syndrome (WS) profile. Overgen = overgeneralization; Nonrhyme +ed = regular generalization of novel items not rhyming with existing irregulars (e.g., *brop–bropped*); Rhyme +ed = regular generalization of novel items rhyming with existing irregulars (*crive–crived*; cf. *drive*); and Rhyme-Irreg = irregularization of novel rhymes (*crive–crove*).

reducing phonological similarity produce identical effects whatever the lexical–semantic code? The results here are particularly interesting. As an illustrative measure, Figure 11 depicts performance on irregular verbs, as phonological similarity was reduced on the one hand and as the effective contribution of lexical–semantics was reduced on the other. As the influence of lexical–semantics weakened, the effect of the phonological manipulation was exaggerated in irregular verb performance. The same was also true of regular verbs. Weakening lexical–semantics also exaggerated the effect of redundancy on irregular verbs, but it reduced the variation caused by redundancy in the behavior of regular verbs.

In short, when at least two main constraints operate, alterations in one constraint may create the conditions under which alterations in the other constraint are exaggerated. That is, in WS, it is possible that a reduced influence of the lexical–semantic system on inflection may lead to the amplification of any atypicalities that exist in the phonological system.

Background Flexibility

Through an examination of a range of network parameters, we sought to evaluate how flexible the model was in accommodating patterns of atypical development. Table 1 shows the results of the various manipulations, classified according to whether each parameter change caused (a) a delay in acquisition, (b) a differential delay to regular and irregular verbs, (c) a final impairment in either verb type, and (d) changes in generalization for each type of novel item. Parameter sets were evaluated against the baseline network. These manipulations were also run in a network that excluded lexical–semantic input, in line with Plunkett and Marchman's (1991) original architecture. When the effects of manipulations differed in the absence of semantics, this is noted in Table 1.

Did any other manipulations reproduce the target pattern of WS data? Three of the 10 parameter sets demonstrated some similarity to the pattern of overall delay and reduced generalization. These

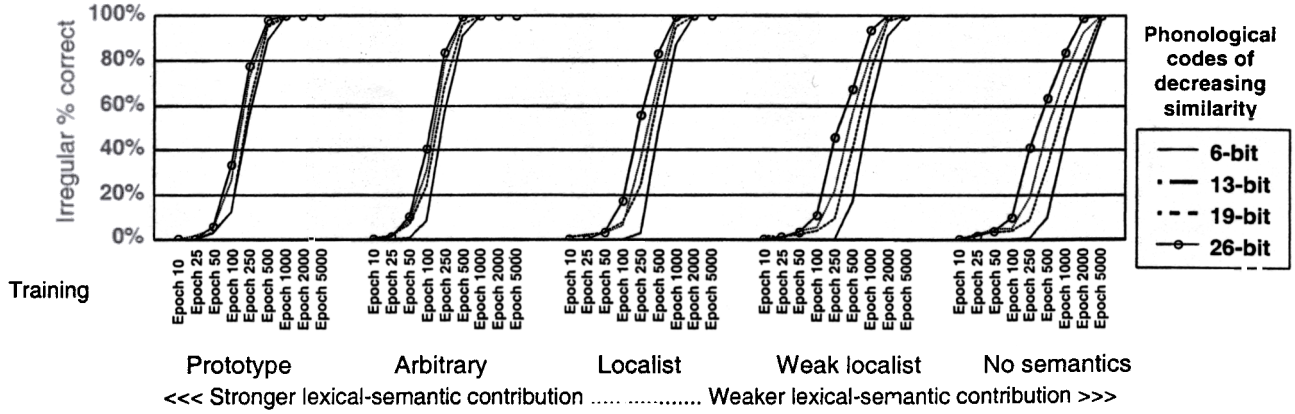

Figure 11. Illustration of the interaction between initial phonological similarity and initial lexical–semantic code on the developmental performance of irregular verbs.

Table 1
Summary of Effects on Network Performance of Each Manipulation Against the Empirical Criteria

Manipulation	Developmental delay?		End-state impairments?	Reduced generalization?			Comments
	Reg.	Irreg.		Nonrhymes + *ed*	Rhymes + *ed*	Rhymes irregularization	
Target data	√	√	X	√	√	√	
Increased HU	X Accelerated	X Accelerated	X	X Increased	√	X Increased	
Decreased HU	√ Reg. more delayed than irreg.	√ NS: irreg. more delayed than reg.	√ Equal for reg. and irreg. (NS: greater for irreg.)	√	√	√	With semantics, end-state deficits do not appear until HU = 10. Without semantics, end-state deficits appear at HU = 20 and are greater for irregular verbs.
Architecture							
2-layer	X Accelerated		X NS: much larger for irreg.	X Increased	X Increased	X No change	Large numbers of overregularization errors.
4-layer (25/25)	√ Reg. more delayed than irreg.	√	X			√	
4-layer (50/50)	√	X Accelerated	X	√	√	X No change	
Fully connected	X Accelerated	X No change	X	X Increased	X Early increase, late no change	X Increased	Direct connections aid acquisition of regulars.
Processing noise	X	X	√ At higher noise levels				With higher processing noise levels (≥ 10%), performance collapses late in training, with irreg. suffering more than reg.
Reduced plasticity in the learning algorithm	/ BP: equal delay	√ SPO irreg. greater delay than reg.	√ SPO irreg. only	√ SPO	X SPO increased	√ SPO	
			X BP (NS: irreg. deficit for both SPO and BP)	X BP no change (NS: BP X increased)	X BP increased	X BP no change	
Increased unit discrimination	√		X	X Increased	√X Early increase, late decrease (NS: x increased)	√ NS: x no change	
Decreased unit discrimination	√ Reg. more delayed than irreg.	√	X			/ NS: x no change	

Note. reg. = regulars; irreg. = irregulars; HU = hidden units; NS = model trained without lexical–semantics; BP = training with backpropagation (reduced plasticity); SPO = training with sigmoid prime offset (intermediate plasticity).

manipulations were a decrease in hidden units (although this decrease needed to be carefully calibrated to avoid a collapse in performance), a four-layer architecture with restricted hidden unit numbers, and reduced discriminability of the processing units. However, in each of these cases, the relative delay for regular verbs tended to be greater than that for irregular verbs.

In terms of outer bounds, when lexical–semantics was available to mediate the effect of parameter changes, its presence protected the performance of irregular verbs. Within the parameter space we examined, we found no selective impairment in irregular verbs at the end of training, so long as lexical–semantic information was available to the network. When lexical–semantic information was absent, final irregular deficits could be generated either by reduced plasticity, decreased hidden units, or the use of a two-layer network.

Table 1 is notable in that, for a given type of atypical response, such as reduced add *–ed* generalization, a number of different parameter manipulations were available that would have this same result (i.e., following down a single column). However, when atypical patterns were considered across a range of responses, such as the six measures illustrated in Table 1, few of the parameter sets produced responses that directly lined up (i.e., comparing whole rows). Indeed, in Table 1, only reduced discriminability and a four-layer architecture with reduced hidden units produce a similar profile, although for the values we used, reduced discriminability had a much larger effect. In short, for this model, a narrowly defined behavioral impairment was not easily traced back to a single underlying atypical constraint. More important, it required consideration of a wider pattern of impairments over several measures before underlying causes could be disambiguated. This idea is explored in more detail by Thomas (2003).

We considered the role of hidden unit numbers in a little more depth. This is because several connectionist models of atypical development have appealed to this parameter as offering explanations of deficits in disorders such as autism and dyslexia. Moreover, Karmiloff-Smith et al. (1997) proposed that an associative system with excessive hidden units might explain their data showing reduced generalization of the phonological regularities of the French gender system by individuals with WS. In our model, we explored the effect of hidden unit numbers on generalization, in the range of 10 to 1,000, with 50 units as our baseline value. A level of 500 units roughly corresponds to 1 hidden unit per training pattern and 1,000 units to 2 per pattern. We evaluated generalization of the add *–ed* regularity to three types of novel pattern: items sharing two phonemes with existing regulars, items sharing two phonemes with existing irregulars, and rare phonotactically illegal items sharing only a single phoneme with any of the verbs in the training set. Figure 12 demonstrates the result, including data for networks with and without lexical–semantic input.

Even at very high levels of hidden units, there was little evidence that generalization of the add *–ed* regularity declined, except on items sharing two phonemes with irregulars. Karmiloff-Smith et al.'s (1997) proposal was thus not supported. Figure 12 also demonstrates that lexical–semantic information facilitates generalization only when hidden unit numbers are small, when such information is helpful in partitioning a restricted representational space. As far as Plunkett and Marchman's (1991) training set is concerned, a surfeit of hidden units did not reduce generalization.

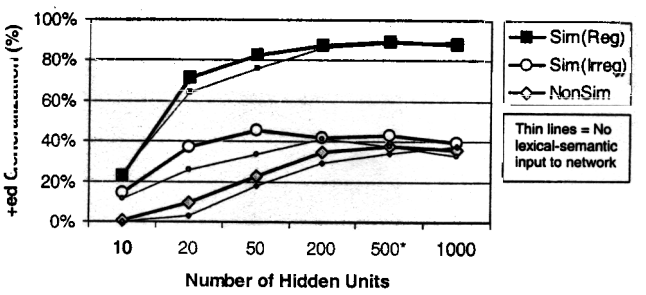

Figure 12. Generalization performance in the network as hidden unit numbers are varied (thick lines). Results are shown for three forms of novel item: Sim(Reg) = novel items sharing two phonemes with an existing regular verb; Sim(Irreg) = novel items sharing two phonemes with an existing irregular verb; NonSim = novel items sharing no more than one phoneme with any verb in the training set. Thin lines show equivalent performance in a network without lexical–semantic input. *500 hidden units approximately equal to 1 unit per verb in the training set. +ed = add *–ed*.

Fitting the Data: How Good Is Best?

We have compared various manipulations, either driven by theoretical hypotheses or by exploration of background flexibility, and found that some are better able than others to show the deficits of the WS data. Before we turn to a discussion of our findings, we address the issue of how good the best fit is.

In the model design, we were interested in the latter phases of performance, because no incrementally expanding vocabulary set was used in training. Figure 13 demonstrates the portion of our baseline condition that is appropriate to the control data from Thomas et al. (2001), along with the effects of the two manipulations that most closely reproduced the pattern of the WS group. The baseline condition best fit the control data by assuming that the performance of the network at 100 epochs was equivalent to the 5–6-year-old age bracket.

The notion of scaling modeling data to developmental trajectories is fraught with difficulties (see Marcus, 1995; Plunkett & Marchman, 1996). For example, both our empirical data and our modeling data used nonlinear scales. However, it is not clear on what basis one chooses the particular nonlinear scale or that we have chosen the correct scale for each set of data. Moreover, from the perspective of the modeling of cognitive and language development, chronological time per se has no causal role. The model is isolated from the context of the general nonlinguistic cognitive system in which it is embedded—developmental events in this wider system may constrain the rate at which the language system can develop. The performance of the model is assessed in exposure to learning instances. It is not clear how the learning instances experienced by the child correspond to the passing of time, or what the appropriate empirical data are that best measure such instances (note the discussion between Marcus, 1995, and Plunkett & Marchman, 1996, about whether Plunkett & Marchman's, 1991, model should be evaluated against parental input to the child or the child's productive vocabulary). As our nonlinear scaling stands, the model manipulations produced smaller delays in past-tense acquisition (~2 years) than those found in a comparison of the WS and control groups (~10 years lag), but calibration of absolute

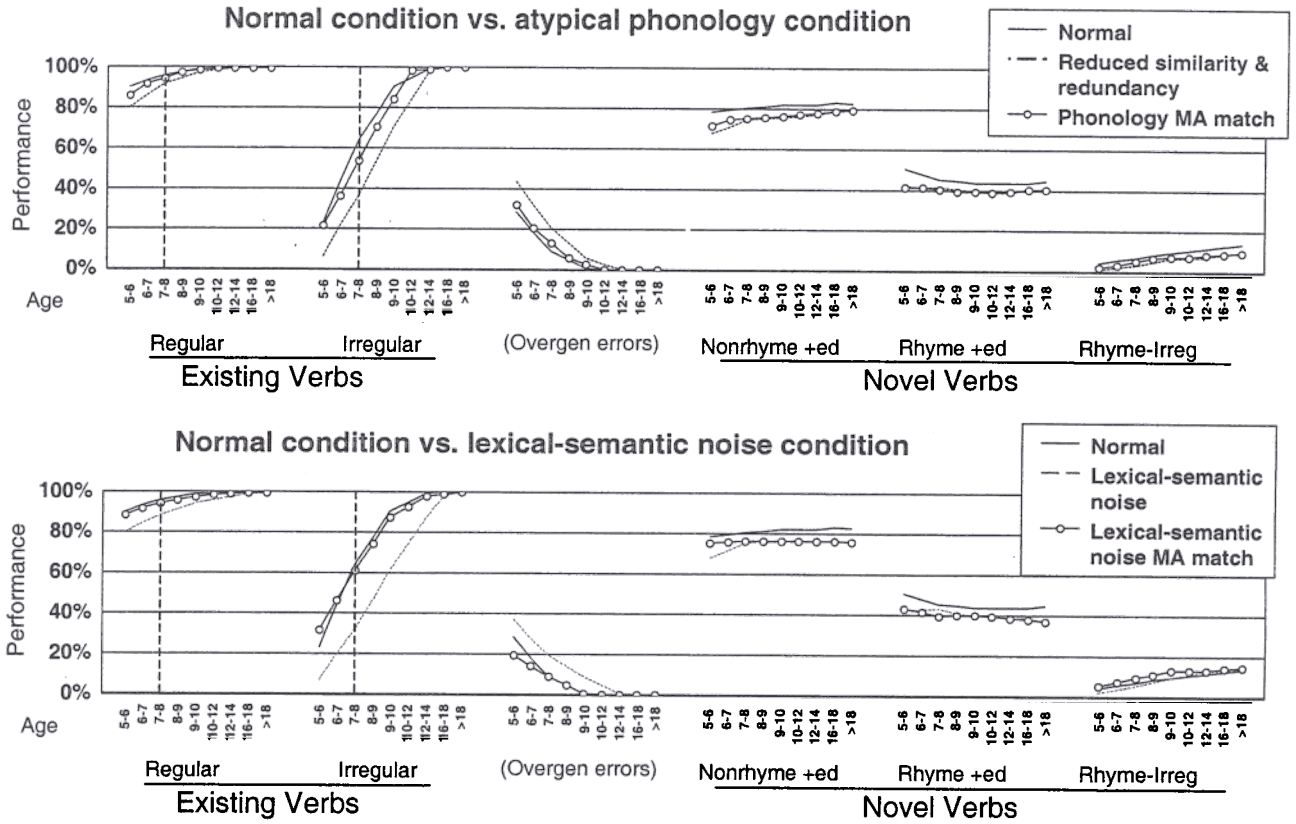

Figure 13. Simulation trajectories plotted for the portion of training relevant to the empirical data. Top: Effect of the phonological manipulation of reduced similarity and redundancy when matched to baseline performance on regular verbs (our implementation of a verbal mental age match). Bottom: Equivalent data for the lexical–semantic noise (integration impairment) condition. Vertical dashed lines show identical points in training for regular and irregular verbs and demonstrate how the apparent deficit for irregular verbs is eliminated by the matching procedure, in line with Thomas et al. (2001). MA = mental age; Overgen = overgeneralization; Nonrhyme +ed = regular generalization of novel items not rhyming with existing irregulars (e.g., *brop–bropped*); Rhyme +ed = regular generalization of novel items rhyming with existing irregulars (*crive–crived*; cf. *drive*); and Rhyme-Irreg = irregularization of novel rhymes (*crive–crove*).

effect size was not an aim for the current model. This task awaits a scaled-up model of inflection.

The scaling we have used produces a good fit to performance on both regular and irregular verb acquisition, although overregularization errors tended to persist in the data somewhat longer than in the model. Regularization of nonrhymes showed a similar profile to the empirical data but was about 15% too low. Irregularization of rhymes produced a similar profile to the control data but failed to replicate the late surge in irregularization rates found in the older age groups. Regularization of rhymes showed the least good fit, with flat generalization rates at around 45%. In contrast, the control group demonstrated an inverted U across this period. Although the model also produced such an inverted-U trajectory (e.g., Figure 3), the model's peak in generalization for these items was lower and occurred earlier in training, such that only the post-peak tail is plotted in Figure 13.

The model's fit to the existing verbs gives confidence that this simple baseline model offers a valid consideration of the effects of initial constraints on the relative difficulty of acquiring regular and irregular inflections. The generalization rates are lower than one would want but still permit consideration of the relative effect on generalization of initial changes. It is these relative effects that we believe will scale to more realistic models of past-tense formation.

In the Method section, we considered the difficulty of comparing simulation data against cross-sectional atypical data, given that age and severity of disorder are confounded in cross-sectional designs. The standard empirical control for level of ability that seeks to disambiguate an atypical profile from general delay is problematic in this context, for empirically it appeals to a measurable system-general delay. For the model, there is no general system. Instead, for a final comparison against the Thomas et al. (2001) data, we sought to control for a version of VMA that was based on the network's level of performance on regular verbs and then to examine how the conditions compared across the other response types. These comparisons, using data from the phonological manipulation of reduced similarity and redundancy and the integration manipulation of lexical–semantic noise, are depicted in Figure 13. These were qualitatively the two best fits to the WS pattern of results.

Figure 13 demonstrates that the relative delay caused by each set of atypical constraints produced an apparently greater deficit for

irregular than regular verbs (a line is drawn at identical stages of training for each verb type to make clear this difference). However, when performance was matched according to regular verbs, in each case the apparent irregular deficit disappeared. Yet, the reduction in generalization in the model remained. For the lexical–semantic noise condition, the reduction was only present in add *–ed* generalization, whereas for the phonological condition, irregularization rates were reduced as well. For the WS data, controlling for VMA left a residual but nonsignificant deficit in rhyme irregularization (Thomas et al., 2001), making it difficult to discriminate between the two best data fits from the model.

Implications of Simulation Results for WS

We begin with a summary of the findings of the model with regard to WS and past tense. Strong theoretical claims about normal language development have been made on the basis of past-tense deficits in WS, namely that a high-level cognitive mechanism in the inflectional system is selectively damaged in a developmental disorder while the rest of that system develops normally. A close examination of available empirical evidence on WS past-tense formation revealed two patterns: (a) an overall delay in both regular and irregular past-tense acquisition, along with a reduction in generalization to novel forms and (b) a less robust pattern of greater delay for irregular past-tense acquisition than for regular. From a review of the literature on language development in WS, we identified five hypotheses proposing how development in this syndrome may differ from the normal case. We used a connectionist model of past-tense development to evaluate implementations of these hypotheses, altering the initial constraints under which development took place and comparing trajectories and outcomes against empirical data.

With regard to the first pattern of empirical data (overall-delay/reduced generalization), the model suggested that this pattern could be produced by a particular anomaly to phonological representations (reduced similarity and redundancy) or by a problem in integrating phonological and lexical–semantic information, whereby noise in the signal from lexical–semantics disrupts the learning of the function relating verb stems and past-tense forms. An examination of the background flexibility of the model suggested that three other, less theoretically motivated parameter manipulations could offer a more approximate fit to the empirical data. These included decreased numbers of hidden units, use of a four-layer architecture with restricted numbers of hidden units in each layer, and a reduction in processing unit discriminability. In these latter cases, there was a tendency for regular verbs to experience a greater delay than irregular verbs, instead of the uniform delay. Simply slowing learning in the network, conversely, produced a uniform delay but no reduction in generalization to novel forms.

With regard to the second pattern of empirical data (greater irregular delay), anomalies in lexical–semantic information (i.e., attenuated activation of this system, anomalous organization, developmentally delayed availability, or slowed weight change in the connections integrating this information) tended to differentially impede irregular verb acquisition. A range of background parameter manipulations generated the same effect, suggesting that irregular verbs were more developmentally vulnerable to nonoptimal conditions in the network.

Three other results were of note. First, Thomas et al. (2001) reported data indicating an increased effect of the semantic dimension of imageability in the irregular past-tense formation of the WS group compared with controls. The simulations suggested that, paradoxically, this is consistent with a reduced influence of lexical–semantic information on inflection in WS, whereby information used by the network to index individual abstract verbs as exceptions from the past-tense add *–ed* regularity becomes too attenuated to prevent overregularization. Second, after Karmiloff-Smith et al. (1997) found reduced levels of generalization in French participants with WS, they speculated that a network learning to predict gender on the basis of phonological regularities might exhibit such a reduction if it had a surfeit of hidden units. The current modeling work did not support this proposal, finding no marked reduction in the generalization of past-tense regularities when hidden unit levels were increased by a factor of 20 over the normal level. Third, the simulations revealed that phonological and lexical–semantic constraints interacted within the developmental system. Specifically, as the contribution of lexical–semantic information was reduced, differences in the initial phonological representations produced more exaggerated developmental effects. Developmental constraints can therefore produce non-additive effects. This has implications for individual variability in developmental disorders: One atypical constraint may not simply generate its own consequences but also exaggerate the effects of (perhaps normal) variation in other constraints. Two points warrant further discussion: the specificity of the successful manipulations and the implications of individual variation.

Manipulations

Two manipulations succeeded in simulating the main WS pattern of overall-delay/reduced generalization: atypical phonological representations and an integration deficit. Further empirical work is required to establish whether one (or both) of these is the correct account. The successful manipulations were also fairly specific: The phonological representations had to have both reduced similarity and redundancy;[4] the integration deficit needed to have a disruptive effect. The specificity of these claims allows more focused future empirical investigations to evaluate their plausibility. In the case of phonology, phoneme discrimination tests would evaluate similarity, whereas resilience under noise would evaluate redundancy. In the case of an integration deficit, studies of language processing in tasks that require the on-line integration of lexical–semantic and phonological constraints would evaluate any disruptions caused in combining sources of information.

[4] The properties of similarity and redundancy could plausibly be linked in the developmental emergence of phonological representations in the language system. Speech-based representations must be developed from acoustic information and under the atypical phonology hypothesis, it is differences in lower level auditory processes that lie at the root of subsequent anomalies in language processes. Oliver et al. (2000) explored the effect of disruptions to the development of self-organizing connectionist systems and found that the specificity of the representations formed as a result of competitive processes was linked to the number of units encoding each of the inputs. Greater competition between processing units in a self-organizing map leads to more specific representations for input items but can lead to more units being deactivated through the process of competition and, therefore, not involved in encoding information. Thus, although the phonological manipulation was fairly precise, it is not an unreasonable one.

Conversely, the alternatively reported WS pattern of differentially delayed irregular past-tense acquisition was more easily simulated. The model is therefore less theoretically constraining as to what may be the actual cause, if this behavioral pattern were to turn out to be robust. However, in practical terms, few data confirm the irregular deficit, and none depict how this may unfold in terms of a developmental trajectory that includes respective regular performance and generalization abilities. Clahsen and Almazan's (1998) data came from 4 participants with WS, split into two groups of 2 for analysis-of-variance purposes. The two groups had mean MAs of 5;6 (using the notation: years;months) and 7;6, respectively (CAs 12;2 and 14;0). If treated as a (scant) developmental profile, these results were consistent with an overall developmental delay in past-tense performance, which is in line with MA for regulars but greater for irregulars, with MA-appropriate regular generalization but reduced irregularization of novel rhymes (e.g., *crive–crove*). In terms of manipulations to the model, the only condition that produced this pattern was a reduction in phonological redundancy alone (see Figure 5: *times1* vs. *normal*). Initial deficits to lexical–semantics could not reproduce this entire pattern: Differentially delayed irregular verb acquisition was typically associated with increased irregularization of novel forms in the network (see Figure 7). Alternatively, Clahsen and colleagues (e.g., Clahsen & Almazan, 1998) have referred to WS as exhibiting an "excessive over-application" (p. 187) of inflectional rules. In behavioral terms, the network condition that best captured this pattern was a two-layered architecture: Without hidden units, the model struggled to acquire irregular verbs, produced many overregularization errors, and exhibited increased regular generalization compared with the baseline model. However, regular verb acquisition was accelerated in this two-layer network, so the profile would not be appropriate to the regular verb delay also shown by individuals with WS.

However, it should be noted that Clahsen and colleagues (e.g., Clahsen & Almazan, 1998) would be reluctant to interpret their findings within a connectionist framework, preferring to relate them to a version of Pinker's (1991, 1994, 1999) DM model involving hierarchically structured lexical representations. We consider the extent to which the simulation results might be extended to that model in the General Discussion section.

Individual Differences

In the Method section, we considered the implications of individual differences exhibited by people with developmental disorders for attempts to capture their behavior in developmental computational models. At that point, we restricted ourselves to modeling only the mean or group performance, while keeping in mind the flexibility of the model. We now turn to consider whether the model could capture individual variation. First, let us establish how much individual variability we might expect in WS and, in particular, in individuals' language profiles.

Generally, there is no doubt that the severity of the WS phenotype can differ across individuals and that the pattern of strengths and weaknesses is blurred by the spectrum of variation that one might expect of a typically developing population. Pezzini, Vicari, Volterra, Milani, and Ossella (1999) recently examined whether in a sample of 18 Italian children with WS (*M* CA = 9;10, range = 4;10–15;3), a single, common neuropsychological profile was discernible. In a comparison of various linguistic and visuospatial tasks, these investigators found a large amount of variability, such that individual neuropsychological profiles often failed to match the group profile. However, they concluded that three behavioral markers were reasonably reliable: (a) an uneven cognitive profile; (b) an impairment in visuospatial construction, indexed by the block-design task; and (c) a dissociation in linguistic skills, with phonological fluency superior to lexical–semantic skills. When Volterra, Longobardi, Pezzini, Vicari, and Antenore (1999) compared the profiles of 10-year-old dizygotic twins, 1 with WS and 1 without, they found that the boy with WS displayed a homogeneous developmental delay in both nonverbal and verbal abilities compared with his typically developing sister. Of the tests examined, he achieved a level of performance similar to his sister only in face recognition, phonological word fluency, and memory for phonologically similar words. Although it is undeniable that individual variation characterizes WS performance (as much as in any atypical group), an imbalance between semantics and phonology seems to be a consistent aspect of language development in this syndrome.

When we focus more specifically on individual variation in performance on past-tense tasks, inspection of the individual scores of the 18 participants in Thomas et al. (2001, Figure 1) demonstrates that performance broadly increased with CA, with several older participants reaching ceiling. Nevertheless, not all of the older participants produced ceiling scores. For example, Participant 20 (CA = 50;11, VMA = 13;0) scored on average 100% on regulars but only 60% on irregulars. Participant 18 (CA = 34;9, VMA = 6;4) scored only 15% on regulars and 7% on irregulars. For these participants with WS, it is doubtful that they will achieve ceiling scores at some later age.

Turning to the model, then, does such arrested development fall within the bounds of the most extreme settings of the atypical parameters? The answer, for the lexical–semantic and phonological manipulations at least, is no. These manipulations, however severe, only slowed but did not prevent ultimate successful acquisition of the domain. For the model to exhibit arrested acquisition of the domain required a combination of manipulations: for example, combining the lexical–semantic or phonological manipulation with a slower learning rate or with reduced processing resources. It is possible that a scaled-up developmental model of inflectional morphology may be more easily disrupted. It is also possible that more severe cases of WS represent combinations of deficits to the cognitive system. These may involve not just alterations to lower level sources of information but deficits in the processing structures available to combine them during development to generate high-level behaviors.

Summary of Implications

The implications of the preceding simulations for research in WS and language development are as follows. First, the modeling results indicate that further work is required on the nature of phonological representations in WS. However, given that standard phonological awareness tasks require a significant meta-cognitive component, more subtle measures may be required (Laing et al., 2001). Second, detailed studies are required that focus on the integration of knowledge sources during on-line language processing in WS (see Karmiloff-Smith et al., 1998). Third, the general-

ization deficit in WS merits further investigation, particularly the conditions under which it arises. For example, Ramscar (2002) has shown that semantic context can influence the inflection of novel forms (presumably by priming the phonological forms of existing verbs consistent with the context). Would this effect hold for WS? Levy and Hermon (in press) have suggested that individuals with WS may show a generalization deficit even when they have a good understanding of the task. To what extent do individuals with WS treat novel terms differently under test conditions? Fourth, if a differential delay in irregular verb inflection were eventually to turn out to be a robust pattern of data in WS (no large study has thus far convincingly demonstrated it), the modeling suggests that a lack of lexical–semantic influence on inflection is a good candidate to explain this differential delay, and detailed empirical work should pursue the nature of this deficit.[5] Finally, one of the main implications from the model is that, rather than an isolated deficient mechanism, multiple atypical constraints are at work in the WS language system. Thus, Thomas et al. (2001) found both reduced generalization and semantic effects in WS performance: Reduced lexical–semantic influence was sufficient in the model to explain the latter but not the former effect. In fact, the model clearly demonstrated that the outer bounds of individual variability had to be accounted for by multiple atypical constraints rather than a single one. In summary, our results indicate that a single atypical developmental constraint is unlikely to explain WS language development.

General Discussion

The computer simulations presented here have focused on how the developmental process itself contributes to end-state deficits in a system with atypical initial constraints. In the introduction, we argued that this neuroconstructivist approach to disorders offers a more realistic view of the origin of developmental deficits than a static explanatory framework borrowed from the methodology of adult cognitive neuropsychology. In the static approach, researchers prefer (when possible) to view a developmentally disordered cognitive system in terms of a normal system that suffers from selective deficits to specific cognitive-level mechanisms. We now wish to amplify the difference between these two approaches.

Before we do so, it is worth pointing out that there are similarities between the approaches. Both types of accounts assume that in many developmental disorders, aspects of normality and abnormality coexist. The crucial difference between the approaches is the developmental stage and the level at which normality and abnormality are postulated to coexist. For the static approach, normality and abnormality coexist in high-level mechanisms in the adult state. For the neuroconstructivist approach, normality and abnormality coexist in lower level computational constraints in the start state. However, the effect of these constraints on end-state adult structures depends crucially on the characteristics of the developmental process that links start state and end state. Here, all too frequently, the static approach explains developmental deficits by reference to a nondevelopmental adult model with an unrealistic developmental account implicitly tacked on.

This can be illustrated in the past-tense domain, if the DM model is used in a static fashion to explain deficits in late childhood onwards in disorders like WS and SLI. Pinker's (1991, 1994, 1999) theory invokes two high-level, domain-specific cognitive mechanisms—one for processing inflectional rules and the other for processing word-specific information—components that are invoked to explain performance in the adult state.[6] The two developmental disorders of WS and SLI are then characterized by appeal to this static model in terms of a double dissociation. Each disorder is argued to correspond to a selective abnormality to one of the high-level mechanisms, whereas the other is "spared" or normal (Pinker, 1999). There is no attempt, here, to identify initial start-state deficits. There is no attempt to chart the developmental consequences of atypical conditions in the start state of either mechanism, before they have taken on their adult, high-level cognitive identities of the components underlying regular and irregular inflection. As a result, there is no attempt to explore the nature and extent of possible compensatory changes between the two mechanisms across development: for example, the possibility that if one mechanism has start-state anomalies, the other may compensate and therefore itself develop atypically. Instead, a simplistic developmental account is merely assumed, whereby somehow, atypical start-state conditions can lead to the abnormal development of one component and the normal development of the other. This developmental assumption, which we have termed elsewhere *residual normality*, turns out to hold only under very specific computational developmental conditions (see Thomas & Karmiloff-Smith, 2002a). That such conditions hold in a given cognitive system cannot be assumed but must be justified empirically.

In contrast, the neuroconstructivist approach sees normality and abnormality as coexisting in the start state (Karmiloff-Smith, 1998). High-level deficits are construed as the outcome of initial lower level deficits plus the application of a developmental process. A consequence of this focus is that the researcher is prompted to question whether seemingly normal behavioral outcomes in the end state may camouflage very subtle deficits, because of the way low-level impairments may eventually impact on the whole developing system. In the past-tense simulations, differences were postulated in the structure of phonological representations, of lexical–semantic representations, or in the computational properties of the systems seeking to generate past-tense forms given phonological and lexical–semantic information. High-level behavioral deficits in, for example, irregular past-tense formation, were the consequence of acquiring the relevant mappings using a computational system with these atypical lower level constraints.

It is worth noting that nothing in the DM model prevents it from being applied to developmental disorders within the neurocon-

[5] Clahsen and Almazan (1998) have a slightly different account of this deficit on the basis of the DM model. In their account, the word-specific information that is unavailable on-line is restricted to the phonological form of irregular past-tense forms and has no semantic content.

[6] For example, here is how Marcus, Pinker, et al. (1992) distinguished Pinker's derivation of the blocking principle (the component of the DM model that coordinates the functioning of the symbolic and associative mechanisms) from a competing proposal of MacWhinney: "The principal difference between MacWhinney's (1978) and Pinker's (1984) expositions is that *Pinker takes the blocking principle as it was explicated and justified by linguists to explain adult knowledge, and simply attributes it to the child,* whereas MacWhinney introduced it as a specific new claim about the child's morphological acquisition system" (p. 16, Footnote 6, italics added).

structivist framework. However, this would entail a serious consideration of the impact of the developmental process. Under a neuroconstructivist treatment of the DM model, start-state deficits (justified by psychological data) would be applied either to the symbolic mechanism, the associative mechanism, or the "epiphany" mechanism that identifies the presence of an inflectional rule in the input. The subsequent developmental trajectory of the model and its end-state behavioral impairments could then be assessed against empirical data, as we have done with the connectionist past-tense model. However, given the potential complexity of the developmental process and the importance of clearly specifying its nature, we have found that computational modeling is almost essential for this type of endeavor. We show in the *Generality and Robustness of Findings* section, below, how this might proceed in regard of the DM model, when we consider computational implementations that bear some similarity to it.

Although we used empirical data from language research in WS to determine which of the model's initial cognitive constraints would be atypical (semantic, phonological, etc.), it is important to realize that these manipulations are not intended to stand as the final explanation for the behavioral deficits in the disorder. Instead, a neuroconstructivist approach prompts the search for an even lower level developmental explanation. We can illustrate this by considering how the neuroconstructivist and static explanatory frameworks would embrace the implications of a genetic cause in a given developmental disorder.

Because the static account views the explanation of a behavioral deficit in terms of selective abnormalities to high-level adult mechanisms, there is a risk that associated genetic anomalies are then construed as the direct cause of the deficit. The affected genes are presumed to be involved in the mechanism's construction during normal development. In this way, theorists using the static approach have recruited disorders like WS and SLI to argue for innate, domain-specific modular structure in the language system, on the basis of different genetic mutations. A theory of this form necessitates a developmental account in which the modules found in the adult cognitive system develop independently under the control of selected genes. However, there is no evidence to support the idea that genes relate directly and solely to high-level cognitive mechanisms in this way (Karmiloff-Smith, Scerif, & Thomas, 2002; Kingsbury & Finlay, 2001; Pallas, 2001; Ragsdale & Grove, 2001).

By contrast, the neuroconstructivist account is predicated on the following recursive equation: High-level deficit = lower level deficit × developmental process. Presented with a genetic disorder, the neuroconstructivist seeks a cascade of these equations, in which each level of atypical structure or function is traced back to a lower level anomaly plus a developmental process. Ultimately, this cascade should reach back to the earliest stage of development in a genetic disorder, that of the embryonic brain.[7]

We believe that such a theoretical approach charts an original approach to future explanations of many developmental disorders. It is a framework in which genes are linked to behavior through multiple levels of cognitive, neurocomputational, and neurobiological modeling of developmental systems. Similarities that may exist between the behavioral deficits found in developmental disorders and acquired disorders may well be informative at some level (see Thomas & Karmiloff-Smith, 2002a), but these should not be employed to motivate the use of the static framework, for this risks the postulation of unrealistically direct links between genes and behavior.

Our illustration of the neuroconstructivist approach to developmental deficits has focused on explaining such deficits in terms of parameter variations to a computational implementation. We encountered a number of theoretical issues implicit in this approach, and we conclude with a brief discussion of three of them. First, we examine how this approach to modeling atypical development fits with attempts to model other forms of cognitive variability. Second, we consider the implication of the multiple causality that was evident from the simulations we presented. Last, we address the extent to which one can draw general conclusions from the results of a particular computational implementation.

Theories of Cognitive Variability

Atypical development represents only one form of cognitive variability. Other forms include the process of development itself, individual differences between normal people of the same age (e.g., intelligence), variations that are due to aging, and day-to-day and even moment-to-moment variations in performance. If one accepts that changes to computational parameters in cognitive models can account for atypical development, one is bound to ask, will such computational parameter manipulations explain other forms of variation? Although this is not the main focus of the present article (see Thomas & Karmiloff-Smith, 2003, for a detailed consideration), it is worth noting some of the issues that a more general approach to cognitive variability would raise. Are alterations to the same computational parameters responsible for each type of variation? Is atypical development simply the tail end of a distribution of normal developmental parameters? Or is it an exaggerated variation that is due to disordered underlying physiology? Is it variation that is perhaps due to very different computational parameters that show no such variation in the normal population? Initial connectionist approaches have at times proposed variations of the same parameters to account for different types of variability, for example with variations in hidden unit numbers being a particular favorite proposed to account for normal development, atypical development, and individual differences. It is unlikely such a "one parameter fits all" approach will be theoretically sustainable (Thomas & Karmiloff-Smith, 2003). A coherent computational picture of cognitive variability is desirable but has yet to emerge. When it does so, we believe that the neuroconstructivist approach elucidated in the present article will be equally applicable in linking different forms of variability to their genetic underpinnings in dynamically evolving systems.

[7] Because causality should not be construed as crossing levels of description, for an account restricted to the psychological level, one should replace in the preceding equation "higher" or "lower level" with "more" or "less abstract" in regard to the representations (see Karmiloff-Smith & Thomas, in press). The relation of brain anomalies to cognitive anomalies is an area of intense debate, as is establishing when it is proper to refer to informational states of the embryonic or neonate brain as "cognitive." These issues are discussed with regard to developmental disorders in Mareschal et al. (in press).

Multiple Causality

In exploring the background flexibility of the model, Table 1 revealed that a given narrow behavioral impairment—for example, a reduction in generalization of the regular past tense—could be generated by a number of different parameter manipulations. Two conclusions follow. When such multiple causality of deficit exists in a model, the model will be less effective in constraining the theoretical inferences that can be drawn from the presence of this particular deficit in a group of individuals or in a syndrome. Second, if the model is valid, individuals exhibiting such a narrow deficit may do so as a result of a range of different underlying impairments. We consider elsewhere the methodological implications of this finding for empirical investigations of behaviorally defined developmental disorders (Thomas, 2003). Here we recognize that connectionist models of atypical development are likely to predict multiple causality of deficits, and we take this as a caution against reading too much into the similarities that different groups show in single behavioral deficits. For example, with regard to the past-tense domain, individual studies have reported deficits in irregular inflection not just in WS but also in children with spinal muscular atrophy (Sieratzki & Woll, 1998) and in children with early and continuously treated phenylketonuria (Badali, Izvorski, Ozawa, Diamond, & Ullman, 1999). The model supports the possibility that the cause could be different in each case.

Computer modeling of development alerts us that different alterations in initial constraints can give rise to similar end-state deficits. It is important to note that computer modeling is one of the few ways in which to explore an opposite possibility: that divergent behavioral outcomes in two disorders may arise from minor differences in otherwise highly similar starting conditions—in other words that, despite quite different end-state behavioral profiles, the causes of two disorders could turn out to be closely related. Although this was raised in previous work as a theoretical possibility (Karmiloff-Smith, 1998), it is only by in-depth modeling that the hypothesis can be fully explored.

We can use the current simulation results to illustrate this point with a hypothetical example. In a past-tense elicitation task, let us say that Disorder A is characterized by an approximately equal developmental delay to regular and irregular past-tense formation but, notably, an increased ability to apply the add *–ed* regularity to novel verbs. Disorder B, conversely, is characterized by a developmental delay that is much more marked for regular verbs than irregular verbs and a very reduced ability to apply the add *–ed* past-tense regularity to novel verbs.[8] Because these patterns were the outcome of particular manipulations, we know the following facts: These apparently contrasting disordered profiles arise from two language subsystems that share the same architecture, representational resources, input and output representations, learning algorithm and learning rate, and levels of processing noise. The two disorders are in fact closely related because they are both caused by an initial alteration to the same system parameter, the discriminability of their constituent processing units. However, the contribution of the developmental process is to push these initially similar start states onto diverging trajectories, producing disorders with contrasting profiles in the end state. The close association of these two hypothetical disorders would not be uncovered by merely focusing on the dissimilarity of the behavioral profiles that they exhibit in the phenotypic outcome. Understanding the two processes of development is crucial.

Generality and Robustness of Findings

The simulations presented in this article amount to the claim that certain developmental deficits in WS can be explained in terms of parameter variations to a particular computational implementation. Such a claim is bound to raise certain objections.

One might object, for example, that the range of developmental deficits explored in the preceding simulations simply indicates that associative networks are fragile learning systems. However, this is not the case here. One should bear in mind that because our aim was to generate atypical trajectories of development, the parameter manipulations we reported were those sufficient to disrupt the course of development, and thus the results may exaggerate the ease with which learning can be disrupted. Nevertheless, under a wide range of manipulations to phonological, lexical–semantic, and architectural parameters, we were able to demonstrate that, surprisingly, the target domain was successfully acquired, albeit at different rates and with different implications for generalization. The associative network was fairly robust in learning the training set. More important, explicit, implemented computational accounts are necessarily more fragile than verbal theories, in that they contain ranges of parameter values within which simulations can capture target data. Specification of these parameter values is what it means to report a successful simulation. Verbal theories often appear robust merely because they lack precise specification. When one specifies theories in greater detail, limitations often become apparent that are simply lost in the mist of vague but plausible verbal theorizing.

One might also object that our thorough exploration of the (developmental) parameter space of a single computation model merely served to shine a 50,000-watt streetlight on a tiny spot of the sidewalk, in the search for the proverbial set of lost keys. How could we guarantee that ours is the right model and that we have found the right manipulation to simulate our target disorder? If ours is not the right model, to what extent would the findings generalize to other possible or existing past-tense models?

These are the right kinds of questions to ask of simulation results, but they must be addressed without undue pessimism about modeling itself. Any of the computational accounts that have been proposed for acquired deficits, for psychiatric disorders, for aging, or for intelligence rely on the assumption that the normal model is the correct one, although no final correct model of the normal system exists in the relevant fields. Yet such models have all generated serious advances in their field. Indeed, no model can prove that it is the right model even if it successfully simulates the data. As has been widely discussed, the role of modeling is more subtle. The process of model construction necessarily involves the cost of simplification, for the benefits of specification, clarification, evaluation of theory viability, generation of testable hypotheses, detailed exploration of the problem domain, and consistency of explanation whereby domains are unified by models that share common computational principles. In the current context, the

[8] Disorder B is not, in fact, that dissimilar to the pattern reported for SLI. See Van der Lely and Ullman (2001).

modeling process led us to consider issues of multiple causality in developmental disorders, the role of individual variability, the computational relation of different forms of cognitive variability, and the possibility of interactions between atypical constraints in developmental systems. Modeling is a tool for theory advancement, often yielding clear empirical predictions.

Of course, it is important to note that we did not choose just any "spot" of the sidewalk on which to shine our 50,000-watt bulb. The spot was based on a prior, thoroughly evaluated computational developmental account of past-tense acquisition, combined with the latest views on the need for differentiated structure in inflectional systems that distinguishes phonological and lexical–semantic inputs (Lavric et al., 2001; Plunkett & Juola, 1999). In addition, as much as possible, we used empirical data to constrain our parameter manipulations, compared multiple manipulations when those data were not sufficiently constraining, and evaluated background flexibility to assess the sensitivity of our results to the particular choices we made for how deficits should be implemented.

Nevertheless, the issue of generality must of course be taken seriously. One of the advantages of siting our exploration of computational accounts of developmental disorders within the past-tense domain is that it permits ready consideration of issues of generality, given the dichotomous nature of the field. We have based our explanations on developmental connectionist models. Would the findings of these simulations generalize to the alternative, DM account of past-tense formation? For example, would reduced phonological similarity and redundancy decrease generalization in that model, or an attenuated lexical–semantic influence differentially delay irregular acquisition? Because the DM model is not computationally implemented, it is impossible to speculate with any confidence. It is here that the disadvantage of underspecified verbal theories becomes apparent.

Given the wider aims and scope of the DM approach, the reluctance of DM theorists to invest time in building implementations of particular domains is understandable. However, as it stands, the DM model is particularly underconstrained when it comes to developmental disorders. As we have suggested, this is because the model cannot specify how, during development, each of its mechanisms might compensate for initial damage to the other.

It is worth exploring this in a little more detail. The DM model could be construed in two ways. First, as we have described it so far, the model might involve a symbolic mechanism, an associative mechanism, and an external epiphany mechanism that identifies the presence of a rule in the input and asks the symbolic mechanism to learn it (Pinker, 1999; see Marcus et al., 1992, pp. 133–137, for details on how this might work). Alternatively, a simpler version of the model might comprise the symbolic and associative learning mechanisms that are exposed to the input without any guidance. The question is, in the case of initial deficiencies in the symbolic mechanism, what is to prevent the associative mechanism from learning all of the past tenses as individual cases, and so producing normal looking end-state behavior? In the case of initial deficiencies in the associative mechanism, what is to stop the symbolic mechanism from learning all the past tenses as rules (either under the direction of the epiphany mechanism in the first version of the model, or on its own in the second version)?

As it has been described in various articles, the DM model certainly appears to have the capacity for such compensatory learning. For example, the symbolic mechanism has the potential to learn multiple rules, invoked to account for the acquisition of the more complicated inflectional paradigms of other languages (Clahsen, 1999, p. 1047). (Indeed the all-rule method was the original approach taken to irregular past-tense formation in linguistics; Chomsky & Halle, 1968.) Moreover, the associative mechanism has the potential to learn regular past tenses as well as exceptions (Pinker, 1999). Thus, Pinker (1999) stated "human memory is not a scarce resource reserved for the incompressible nuggets that cannot be generated by rules" (p. 138).

Of course, one could specify initial constraints that would make developmental compensation more difficult: for example, if the epiphany mechanism functioned independently of the DM system, so that the symbolic mechanism was guided to learn the same inflectional rules irrespective of the performance of the associative memory.[9] Or if there were a particular limit on the number of rules that could be learned. Or if there were a particular limit on the number of forms that could be memorized. However, until such developmental constraints have been clearly specified, the implications for initial deficiencies in one mechanism on what the other may be able to acquire remain pure speculation.

Existing implementations, alternatively, provide more solid ground for evaluating the generality of the claims made from our simulations. As we indicated in the introduction, DM theorists have identified some current computational implementations that bear similarities to the DM account, in that they involve a memory device with a separate rule-like device (although in most cases, the rule device just serves to copy the verb stem to the output). In the next four paragraphs, we consider how our findings would generalize to four DM-consistent computational implementations. These implementations would correspond to the alternative, simpler version of the DM model identified above, without the epiphany mechanism. From these four cases, a consistent picture emerges.

Westermann and Goebel's (1995) model of German verb inflection appears on first impression to have a lexical memory and a copy function to aid in regular inflection. Closer inspection reveals that the production part of the network in fact generates the inflected form from a combination of the phonological representation of the stem and a localist representation of individual word identity. This structure is therefore actually very similar to the architecture we have explored in our model, albeit implemented in a recurrent system. Westermann (1995) found that removal of or restrictions to the lexical memory differentially delayed irregular inflection but also delayed regular inflection, because the initially intact components of the network had to learn both inflection types. It is less obvious what implications would ensue from changes to the phonological representations: This may affect development of the lexical component (a phonology-driven, self-organizing map), and it may affect on-line generalization, because in the model novel forms that rhyme with existing verbs cause lexical activation of the relevant existing verb. Reduced similarity may reduce the lexical activation for novel forms and, therefore,

[9] See Thomas and Karmiloff-Smith (2002a) for a discussion of guided specialization as a computational constraint that produces selective end-state deficits along with residual normality in a developmental disorder.

alter generalization properties (Westermann, personal communication, January 12, 2002).

Second, Westermann (1998) used a constructivist architecture in which, again, one part of the model (direct input–output connections) came to specialize in regular inflection, whereas another parallel processing route, a hidden layer of increasing size, came to specialize in irregular inflection. Omission or developmental restriction of the growing hidden layer caused a delay for irregular inflection and a lesser delay for regular inflection (Westermann, personal communication, January 12, 2002).

Our third example, an inflectional model within the ACT-R paradigm proposed by Taatgen and Anderson (2002; Taatgen, 2001), has similarities to the DM account in that it combines rule-based processing with a memory for instances. Thus, it shares the hybrid representational commitments of the DM model. Taatgen and Anderson's model includes no representation of phonological structure, so manipulations to phonological similarity cannot be considered at this stage. The implications of differences to the model's lexical memory cannot be estimated, because in part the memory and rule component work in series rather than in parallel (for example, *retrieve past tense* is a rule that operates on forms stored in lexical memory). Under normal conditions, the fully trained end-state model tends to rely on this retrieve-stored-form rule to inflect most verbs. Under atypical conditions of lexical memory, it is unclear what compensatory strategies would be available to the system or how these might differentially affect regular/irregular verbs. As it stands, for example, it is not obvious that a regular rule would emerge *at all* if lexical memory were attenuated, because emergence of the default add *–ed* rule is contingent on some prior history of forming new past tenses by analogy to random examples retrieved from lexical memory. The model's initial default is simply to reproduce an uninflected stem, and with a failure of lexical memory, it might just persist in this behavior.

Finally, Hare et al. (1995) proposed a model that combines two components: a three-layer connectionist network that maps between stem and inflected form, and a copy function that supplies the stem to the output layer. The second component is included to offer an improvement in default generalization. From Marcus's (2001) perspective as a DM theorist, these components correspond to two qualitatively different mechanisms operating in parallel, one as a lexical memory, the other as a rule implementing "Copy X." What might happen under developmental failure of either component? Failure of the lexical memory component would restrict the model to outputting an uninflected stem (just as in the preceding model, by coincidence) because only the copy function would remain. Changes to phonology would most likely reduce generalization of the add *–ed* regularity in line with our simulations, because the lexical memory component of the Hare et al. model must still specify the conditions under which the *–ed* suffix is added to the stem. The copy function ensures that novel forms incorporate the stem in the output in which lexical memory does not specify an irregular, but the rule component itself is unable to generate the suffix.

In summary, it appears that when components of inflectional models help supply or process word-specific information, developmental anomalies to these components are likely to produce a greater delay for irregular inflection, although regulars may also suffer through compensatory processes. It also seems likely that changes in phonology will have implications for generalization, although the exact ramifications are hard to predict without actual simulation. In short, two of the major effects we report seem likely to generalize to more DM-consistent models. Of course, these predictions would need to be verified by simulation: If one could fully anticipate the behavior of complex learning systems, there would be no need for simulation at all.

This brief review of DM-consistent models suggests that there is some generality of our simulation results and, thus, that our conclusions are not based on particular implementation details. Nevertheless, none of the four preceding models has been thoroughly examined within the developmental domain, and they stand here as a proxy for the currently underspecified DM account (none is an explicit implementation of the DM account endorsed by those theorists). We have embarked on this consideration of generality to illustrate that, if one chooses to use simulations to formulate explanations of developmental deficits, one must go beyond simulation results to distinguish general principles from any particular implementation details. The theoretically dichotomous past-tense field provides an ideal forum for this endeavor. Crucially, however, it highlights the importance of implementation to generate a focused debate in the field. In terms of theoretical approaches to developmental disorders, currently it is most important to improve our characterization of the process of development itself, and, as we have illustrated, computational implementation is a faithful servant to that goal.

Conclusion

An in-depth consideration of a developmental computational model of past-tense formation in WS made it possible to clarify important theoretical debates in that domain. It then allowed us to evaluate more widely the advantages, disadvantages, and hidden assumptions of using developmental computational models to explain behavioral deficits in developmental disorders. We conclude that, provided their assumptions are well understood, computational models of development (here represented by connectionist networks) have great potential to aid in the understanding of deficits in developmental disorders because they focus attention on the developmental process itself as a crucial causal factor. This contrasts with previous static approaches seeking to characterize such deficits in terms of selective damage to high-level components, analogous to cases of adult brain damage. Disordered systems are those that develop under atypical low-level constraints. Our contention is that computational models of learning are an excellent tool to study atypical processes of development in complex systems.

References

Anderson, V., Northam, E., Hendy, J., & Wrennall, J. (2001). *Developmental neuropsychology: A clinical approach.* Hove, England: Psychology Press.

Badali, S., Izvorski, R., Ozawa, K., Diamond, A., & Ullman, M. T. (1999, April). *Phenylketonuria as a model for investigating the role of dorsolateral prefrontal cortex in language.* Paper presented at the 6th Annual Meeting of the Cognitive Neuroscience Society, Washington, DC.

Baron-Cohen, S. (1998). Modularity in developmental cognitive neuropsychology: Evidence from autism and Gilles de la Tourette syndrome. In J. A. Burack, R. M. Hodapp, & E. Zigler (Eds.), *Handbook of mental*

retardation and development (pp. 334–348). Cambridge, England: Cambridge University Press.

Bates, E. (Ed.). (1991). Cross-linguistic studies of aphasia [Special issue]. *Brain and Language, 41.*

Bellugi, U., Marks, S., Bihrle, A., & Sabo, H. (1988). Dissociation between language and cognitive functions in Williams syndrome. In D. Bishop & K. Mogford (Eds.), *Language development in exceptional circumstances* (pp. 177–189). London: Churchill Livingstone.

Bellugi, U., Wang, P., & Jernigan, T. L. (1994). Williams syndrome: An unusual neuropsychological profile. In S. Broman & J. Grafman (Eds.), *Atypical cognitive deficits in developmental disorders: Implications for brain function* (pp. 23–56). Hillsdale, NJ: Erlbaum.

Berko, J. (1958). The child's learning of English morphology. *Word, 14,* 150–177.

Bird, H., Lambon Ralph, M. A., Seidenberg, M. S., McClelland, J. L., & Patterson, K. (2003). Deficits in phonology and past-tense morphology: What's the connection? *Journal of Memory and Language, 48,* 502–526.

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

Bishop, D. V. M. (1999, December 17). An innate basis for language? *Science, 286,* 2283–2284.

Bishop, D. V. M., North, T., & Donlan, C. (1995). Genetic basis of specific language impairment. *Developmental Medicine & Child Neurology, 37,* 56–71.

Böhning, M., Campbell, R., & Karmiloff-Smith, A. (2002). Audiovisual speech perception in Williams syndrome. *Neuropsychologia, 40,* 1396–1406.

Bromberg, H., Ullman, M., Coppola, M., Marcus, G., Kelley, K., & Levine, K. (1994, July). *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.

Brown, G. D. A. (1997). Connectionism, phonology, reading, and regularity in developmental dyslexia. *Brain and Language, 59,* 207–235.

Bullinaria, J. A. (1997). Modelling reading, spelling, and past tense learning with artificial neural networks. *Brain and Language, 59,* 236–266.

Cáceres, J. A., Heinze, E. G., & Méndez, M. S. (1999). *Preliminary evaluation of some cognitive, social, linguistic, and personality characteristics in a Spanish sample with Williams syndrome.* Unpublished manuscript.

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

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper Row.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioural and Brain Sciences, 22,* 991–1060.

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

Clahsen, H., & Almazan, M. (2001). Compounding and inflection in language impairment: Evidence from Williams syndrome (and SLI). *Lingua, 111,* 729–757.

Clahsen, H., & Temple, C. (2003). Words and rules in Williams syndrome. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations* (pp. 323–352). Hillsdale, NJ: Erlbaum.

Cohen, I. L. (1994). An artificial neural network analogue of learning in autism. *Biological Psychiatry, 36,* 5–20.

Cohen, I. L. (1998). Neural network analysis of learning in autism. In D. Stein & J. Ludick (Eds.), *Neural networks and psychopathology* (pp. 274–315). Cambridge, England: Cambridge University Press.

Cottrell, G. W., & Plunkett, K. (1994). Acquiring the mapping from meaning to sounds. *Connection Science, 6,* 379–412.

Donnai, D., & Karmiloff-Smith, A. (2000). Williams syndrome: From genotype through to the cognitive phenotype. *American Journal of Medical Genetics: Seminars in Medical Genetics, 97,* 164–171.

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.

Ervin, S. M. (1964). Imitation and structural change in children's language. In E. H. Lenneberg (Ed.), *New directions in the study of language* (pp. 163–189). Cambridge, MA: MIT Press.

Fahlman, S. E. (1988). *An empirical study of learning speed in back-propagation networks* (Tech. Rep. CMU-CS-88-162). Pittsburgh, PA: Carnegie-Mellon University, Computer Science Department.

Frawley, W. (2002). Control and cross-domain mental computation: Evidence from language breakdown. *Computational Intelligence, 18,* 1–28.

Fromkin, V., & Rodman, R. (1988). *An introduction to language* (4th ed.). London: Holt, Rinehart & Winston.

Goebel, R., & Indefrey, P. (2000). A recurrent network with short term memory capacity learning the German –s plural. In J. Murre & P. Broeder (Eds.), *Models of language acquisition* (pp. 177–200). New York: Oxford University Press.

Grant, J., Karmiloff-Smith, A., Berthoud, I., & Christophe, A. (1996). Is the language of people with Williams syndrome mere mimicry? Phonological short-term memory in a foreign language. *Cahiers de Psychologie Cognitive, 15,* 615–628.

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

Gustafsson, L. (1997). Inadequate cortical feature maps: A neural circuit theory of autism. *Biological Psychiatry, 42,* 1138–1147.

Hahn, U., & Nakisa, R. C. (2000). German inflection: Single or dual route? *Cognitive Psychology, 41,* 313–360.

Hare, M., Elman, J. L., & Daugherty, K. G. (1995). Default generalisation in connectionist networks. *Language and Cognitive Processes, 10,* 601–630.

Harm, M., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106,* 491–528.

Hinton, G. (1989). Connectionist learning procedures. *Artificial Intelligence, 40,* 185–234.

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Volume 1. Foundations* (pp. 282–317). Cambridge, MA: MIT Press.

Hoeffner, J. (1992). Are rules a thing of the past? The acquisition of verbal morphology by an attractor network. In *Proceedings of the 14th Annual Meeting of the Cognitive Science Society* (pp. 861–866). Hillsdale, NJ: Erlbaum.

Hoeffner, J. H., & McClelland, J. L. (1993). Can a perceptual processing deficit explain the impairment of inflectional morphology in developmental dysphasia? A computational investigation. In E. V. Clark (Ed.), *Proceedings of the 25th Child Language Research Forum.* Palo Alto, CA: Stanford University Press.

Howlin, P., Davies, M., & Udwin, O. (1998a). Cognitive functioning in adults with Williams syndrome. *Journal of Child Psychology and Psychiatry, 39,* 183–189.

Howlin, P., Davies, M., & Udwin, O. (1998b). Syndrome specific characteristics in Williams syndrome: To what extent do early behavioural patterns persist into adult life? *Journal of Applied Research in Intellectual Disabilities, 11,* 207–226.

Jarrold, C., Phillips, C. E., Baddeley, A. D., Grant, J., & Karmiloff-Smith, A. (2001, July). *Comprehension of spatial and non-spatial language in Williams syndrome.* Paper presented at the Experimental Psychology Society Meeting, Manchester, NH.

Joanisse, M. F. (2000). *Connectionist phonology.* Unpublished doctoral dissertation, University of Southern California.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology following brain injury: A connectionist model. *Proceedings of the National Academy of Sciences, USA, 96,* 7592–7597.

Johnson, M. (1999). Cortical plasticity in normal and abnormal cognitive development: Evidence and working hypothesis. *Development and Psychopathology, 11,* 419–437.

Johnson, S., & Carey, S. (1998). Knowledge enrichment and conceptual change in folk biology: Evidence from Williams syndrome. *Cognitive Psychology, 37,* 156–184.

Jones, W., Bellugi, U., Lai, Z., Chiles, M., Reilly, J., Lincoln, A., & Adolphs, R. (2000). Hypersociability in Williams syndrome. *Journal of Cognitive Neuroscience, 12*(Suppl.), 30–46.

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., Klima, E., Bellugi, U., Grant, J., & Baron-Cohen, S. (1995). Is there a social module? Language, face processing, and theory of mind in individuals with Williams syndrome. *Journal of Cognitive Neuroscience, 7,* 196–208.

Karmiloff-Smith, A., Scerif, G., & Thomas, M. S. C. (2002). Different approaches to relating genotype to phenotype in developmental disorders. *Developmental Psychobiology, 40,* 311–322.

Karmiloff-Smith, A., & Thomas, M. S. C. (in press). What can developmental disorders tell us about the neurocomputational constraints that shape development? The case of Williams syndrome. *Development and Psychopathology.*

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.

Kingsbury, M. A., & Finlay, B. L. (2001). The cortex in multidimensional space: Where do cortical areas come from? *Developmental Science, 2,* 125–143.

Kraus, M., & Penke, M. (2000). *Inflectional morphology in German Williams syndrome.* Unpublished manuscript, University of Duesseldorf, Duesseldorf, Germany.

Kuczaj, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behaviour, 16,* 589–600.

Laing, E., Butterworth, G., Ansari, D., Gsödl, M., Longhi, E., Panagiotaki, G., et al. (2002). Atypical development of language and social communication in toddlers with Williams syndrome. *Developmental Science, 5,* 233–246.

Laing, E., Hulme, C., Grant, J., & Karmiloff-Smith, A. (2001). Learning to read in Williams syndrome: Looking beneath the surface of atypical reading development. *Journal of Child Psychology and Psychiatry, 42,* 729–739.

Lavric, A., Pizzagalli, D., Forstmeier, S., & Rippon, G. (2001). Mapping dissociations in verb morphology. *Trends in Cognitive Sciences, 5,* 301–308.

Leonard, L., Bortolini, U., Caselli, M. C., McGregor, K., & Sabbadini, L. (1992). Morphological deficits in children with specific language impairment: The status of features in the underlying grammar. *Language Acquisition, 2,* 151–179.

Levy, Y., & Hermon, S. (in press). Morphology in children with Williams syndrome: Evidence from Hebrew. *Developmental Neuropsychology.*

MacWhinney, B. (1978). Processing a first language: The acquisition of morphophonology. *Monographs of the Society for Research in Child Development, 43*(1–2, Serial No. 174).

MacWhinney, B., & Bates, E. (1989). *The crosslinguistic study of sentence processing.* New York: Cambridge University Press.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition, 40,* 121–157.

Majerus, S., Palmisano, I., van der Linden, M., Barisnikov, K., & Poncelet, M. (2001). An investigation of phonological processing in Williams syndrome. *Journal of the International Society, 7,* 153.

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21,* 339–366.

Marcus, G. F. (1995). The acquisition of the English past tense in children and multilayered connectionist networks. *Cognition, 56,* 271–279.

Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science.* Cambridge, MA: MIT Press.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., Woest, A., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology, 29,* 189–256.

Marcus, G., Pinker, S., Ullman, M., Hollander, J., Rosen, T., & Xu, F. (1992). Overregularisation in language acquisition. *Monographs of the Society for Research in Child Development, 57*(Serial No. 228).

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

Mareschal, D., & Thomas, M. S. C. (2001). Self-organisation in normal and abnormal cognitive development. In A. F. Kalverboer & A. Gramsbergen (Eds.), *Handbook of brain and behaviour in human development* (pp. 743–766). Dordrecht, the Netherlands: Kluwer.

Massaro, D. W. (1988). Some criticisms of connectionist models of human performance. *Journal of Memory and Language, 27,* 213–234.

McCarthy, R. A., & Warrington, E. K. (1990). *Cognitive neuropsychology.* New York: Academic Press.

Mervis, C., & Bertrand, J. (1997). Developmental relations between cognition and language: Evidence from Williams Syndrome. In L. B. Adamson & M. A. Romski (Eds.), *Research on communication and language disorders: Contributions to theories of language development* (pp. 75–106). New York: Brookes.

Mervis, C. B., Morris, C. A., Bertrand, J., & Robinson, B. F. (1999). William syndrome: Findings from an integrated program of research. In H. Tager-Flusberg (Ed.), *Neurodevelopmental disorders* (pp. 65–110). Cambridge, MA: MIT Press.

Moore, M., & Johnston, J. (1993). Expressions of past time by normal and language-impaired children. *Journal of Communication Disorders, 28,* 57–72.

Nazzi, T., Paterson, S., & Karmiloff-Smith, A. (2003). Early word segmentation by infants and toddlers with Williams syndrome. *Infancy, 4,* 251–271.

Neville, H. J. (1991). Neurobiology of cognitive and language processing: Effects of early experience. In K. R. Gibson & A. C. Peterson (Eds.), *Brain maturation and cognitive development: Comparative and cross-cultural perspectives. Foundation of human behaviour* (pp. 355–380). New York: Aldine de Gruyter.

Neville, H. J., Holcomb, P. J., & Mills, D. M. (1989). Auditory, sensory and language processing in Williams syndrome: An ERP study. *Journal of Clinical and Experimental Neuropsychology, 11,* 52.

Neville, H. J., Mills, D. L., & Bellugi, U. (1994). Effects of altered auditory sensitivity and age of language acquisition on the development of language-relevant neural systems: Preliminary studies of Williams syndrome. In S. Broman & J. Grafman (Eds.), *Atypical cognitive deficits in developmental disorders: Implications for brain function* (pp. 67–83). Hillsdale, NJ: Erlbaum.

Newfield, M. U., & Schlanger, B. B. (1968). The acquisition of English

morphology by normal and educable mentally retarded children. *Journal of Speech and Hearing Research, 11,* 693–706.

Oetting, J., & Horohov, J. (1997). Past tense marking by children with and without specific language impairment. *Journal of Speech and Hearing Research, 40,* 62–74.

Oliver, A., Johnson, M. H., Karmiloff-Smith, A., & Pennington, B. (2000). Deviations in the emergence of representations: A neuroconstructivist framework for analysing developmental disorders. *Developmental Science, 3,* 1–23.

Pallas, S. L. (2001). Intrinsic and extrinsic factors that shape neocortical specification. *Trends in Neurosciences, 24,* 417–423.

Paterson, S. J. (2000). *The development of language and number understanding in Williams syndrome and Down's syndrome: Evidence from the infant and mature phenotypes.* Unpublished doctoral thesis, University College London, London.

Pezzini, G., Vicari, S., Volterra, V., Milani, L., & Ossella, M. T. (1999). Children with Williams syndrome: Is there a single neuropsychological profile? *Developmental Neuropsychology, 15,* 141–155.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Pinker, S. (1991, August 2). Rules of language. *Science, 253,* 530–535.

Pinker, S. (1994). *The language instinct.* London: Penguin Books.

Pinker, S. (1999). *Words and rules.* London: Weidenfeld & Nicolson.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28,* 73–193.

Plaut, D. C. (1995a). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 17,* 291–231.

Plaut, D. C. (1995b). Semantic and associative priming in a distributed attractor network. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 37–42). Hillsdale, NJ: Erlbaum.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377–500.

Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science, 23,* 463–490.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition, 38,* 1–60.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48,* 21–69.

Plunkett, K., & Marchman, V. (1996). Learning from a connectionist model of the English past tense. *Cognition, 61,* 299–308.

Plunkett, K., Sinha, C., Møller, M. F., & Strandsby, O. (1992). Vocabulary growth in children and a connectionist net. *Connection Science, 4,* 293–312.

Ragsdale, C. W., & Grove, E. A. (2001). Patterning in the mammalian cerebral cortex. *Current Opinion in Neurobiology, 11,* 50–58.

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense doesn't require a rule. *Cognitive Psychology, 45,* 45–94.

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 behaviour* (pp. 367–392). New York: Cambridge University Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Volume 1. Foundations* (pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Volume 2. Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 452–477.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, England: Cambridge University Press.

Sieratzki, J. S., & Woll, B. (1998). Toddling into language: Precocious language development in motor-impaired children with spinal muscular atrophy. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the 22nd Annual Boston University Conference on Language Development* (Vol. 2, pp. 684–694). Somerville, MA: Cascadilla Press.

Simon, T. J., & Halford, G. S. (1995). *Developing cognitive competence: New approaches to process modeling.* Hillsdale, NJ: Erlbaum.

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.

Stevens, T., & Karmiloff-Smith, A. (1997). Word learning in a special population: Do individuals with Williams syndrome obey lexical constraints? *Journal of Child Language, 24,* 737–765.

Taatgen, N. A. (2001). Extending the past-tense debate: A model of the German plural. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 1018–1023). Mahwah, NJ: Erlbaum.

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say "broke"? A model of learning the past tense without feedback. *Cognition, 86,* 123–155.

Temple, C. (1997). *Developmental cognitive neuropsychology.* Hove, England: Psychology Press.

Temple, C., Almazan, M., & Sherwood, S. (2002). Lexical skills in Williams syndrome: A cognitive neuropsychological analysis. *Journal of Neurolinguistics, 15,* 463–495.

Thomas, M. S. C. (2000). Neuroconstructivism's promise. *Developmental Science, 3,* 35–37.

Thomas, M. S. C. (2002). Development as a cause in developmental disorders. *Computational Intelligence, 18,* 50–54.

Thomas, M. S. C. (2003). Multiple causality in developmental disorders: Methodological implications from computational modelling. *Developmental Science, 6,* 537–556.

Thomas, M. S. C., Dockrell, J., Messer, D., Parmigiani, C., Ansari, D., & Karmiloff-Smith, A. (2002). *Naming in Williams syndrome.* Manuscript submitted for publication.

Thomas, M. S. C., Grant, J., Barham, Z., Gsödl, M., Laing, E., 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. (2002a). Are developmental disorders like cases of adult brain damage? Implications from connectionist modeling. *Behavioral and Brain Sciences, 25,* 727–788.

Thomas, M. S. C., & Karmiloff-Smith, A. (2002b). Modelling typical and atypical cognitive development. In U. Goswami (Ed.), *Handbook of childhood development* (pp. 575–599). Oxford, England: Blackwell.

Thomas, M. S. C., & Karmiloff-Smith, A. (2003). Connectionist models of development, developmental disorders and individual differences. In R. J. Sternberg, J. Lautrey, & T. Lubart (Eds.), *Models of intelligence: International perspectives* (pp. 133–150). Washington, DC: American Psychological Association.

Thomas, M. S. C., van Duuren, M., Ansari, D., Parmigiani, C., & Karmiloff-Smith, A. (2002). *The development of semantic categories and metaphor comprehension in Williams syndrome.* Manuscript in preparation.

Tyler, L. K., de Mornay Davies, P., Anokhina, R., Longworth, C., Randall, B., & Marslen-Wilson, W. D. (2002). Dissociations in processing past tense morphology: Neuropathology and behavioural studies. *Journal of Cognitive Neuroscience, 14,* 79–94.

Tyler, L. K., Randall, B., & Marslen-Wilson, W. D. (2002). Phonology and neuropsychology of the English past tense. *Neuropsychologia, 40,* 1154–1166.

Udwin, O., & Yule, W. (1991). A cognitive and behavioural phenotype in Williams syndrome. *Journal of Clinical and Experimental Neuropsychology, 13,* 232–244.

Ullman, M. T. (in press). Evidence that lexical memory is part of temporal lobe declarative memory, and that grammatical rules are processed by the frontal/basal-ganglia procedural system. *Brain and Language.*

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience, 9,* 266–276.

Ullman, M. T., & Gopnik, M. (1999). Inflectional morphology in a family with inherited specific language impairment. *Applied Psycholinguistics, 20,* 51–117.

Ullman, M. T., Izvorski, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (in press). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language.*

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.

Vicari, S., Brizzolara, D., Carlesimo, G., Pezzini, G., & Volterra, V. (1996). Memory abilities in children with Williams syndrome. *Cortex, 32,* 503–514.

Vicari, S., Carlesimo, G., Brizzolara, D., & Pezzini, G. (1996). Short-term memory in children with Williams syndrome: A reduced contribution of lexical-semantic knowledge to word span. *Neuropsychologia, 34,* 919–925.

Volterra, V., Capirci, O., & Caselli, M. C. (2001). What atypical populations can reveal about language development: The contrast between deafness and Williams syndrome. *Language and Cognitive Processes, 16,* 219–239.

Volterra, V., Capirci, O., Pezzini, G., Sabbadini, L., & Vicari, S. (1996). Linguistic abilities in Italian children with Williams syndrome. *Cortex, 32,* 663–677.

Volterra, V., Longobardi, E., Pezzini, G., Vicari, S., & Antenore, C. (1999). Visuo-spatial and linguistic abilities in a twin with Williams syndrome. *Journal of Intellectual Disability Research, 43,* 294–305.

Westermann, G. (1995). *Connectionist rule processing: A neural network that learns the German participle.* Unpublished master's thesis, Technische Universität, Braunschweig, Germany.

Westermann, G. (1998). Emergent modularity and U-shaped learning in a constructivist neural network learning the English past tense. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th Annual Meeting of the Cognitive Science Society* (pp. 1130–1135). Hillsdale, NJ: Erlbaum.

Westermann, G., & Goebel, R. (1995). Connectionist rules of language. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 236–241). Hillsdale, NJ: Erlbaum.

Xu, F., & Pinker, S. (1995). Weird past tense forms. *Journal of Child Language, 22,* 531–556.

Zorzi, M., Houghton, G., & Butterworth, B. (1998a). The development of spelling-sound relationships in a model of phonological reading. *Language and Cognitive Processes, 13,* 337–372.

Zorzi, M., Houghton, G., & Butterworth, B. (1998b). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1131–1161.

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

Received May 19, 2000
Revision received February 25, 2002
Accepted August 9, 2002 ■