J. Child Lang. **26** (1999), 217–260. Printed in the United Kingdom
© 1999 Cambridge University Press

REVIEW ARTICLE AND DISCUSSION

Rethinking innateness

MATTHEW RISPOLI
Arizona State University

Review essay on: ELMAN, J., BATES, E., JOHNSON, M., KARMILOFF-SMITH, A., PARISI, D. & PLUNKETT, K. *Rethinking innateness: a connectionist perspective on development*. Cambridge, MA: MIT Press (1996). Pp. 447.

I believe that the field of developmental psycholinguistics suffers from two major weaknesses. The first is its impressionistic and inexact formulations. The second is its divisive polarizations. One can see the reasons for the first weakness. Developmental psycholinguistics is only about 30 years old (ignoring diary studies which preceded the linguistic and cognitive surge of the sixties). But speculation and hypothesizing on the basis of relatively little data and passing acquaintance with phenomena has reached the level of customary 'business as usual'. We are skilful at hypothesis construction, yet we are regretfully delinquent at formulating clear tests of our hypotheses. We are fond of conjecture about causal relationships, but our empirical tests progress no further than weak forms of correlation.

With regard to our second major weakness, our knack for polarizing opinion regarding chimerical questions such as the innateness of language can also be understood. After all, are we not following the classical dialectic model of thesis, antithesis and eventual synthesis? I think this is an idealized view of ourselves. In fact, we are driven by hunch and bias far more often than we would like to admit. Following hunches may be a real sign of creativity and vitality in our thinking. However, polarization driven by biases is ultimately detrimental. At some point we must disentangle ourselves from customary dialogue and transcend our deeply rutted patterns of thought.

When I began *Rethinking innateness*, I had hoped that the book might help us overcome these two weaknesses. I believe that it contributes positively to the goal of increasing the precision of our hypotheses and their empirical substantiation. At the same time, I am afraid that it will have a negative impact by aggravating the degree of polarization in our field.

Before proceeding to a discussion of the virtues and weaknesses of this book, a brief overview is in order. The book is divided into seven chapters: (1) New perspectives on development; (2) Why connectionism?; (3) Ontogenetic development: a connectionist synthesis; (4) The shape of change; (5) Brain development; (6) Interactions, all the way down, and (7) Rethinking

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

innateness. Each chapter contributes to the construction of a central thesis that developmental processes are best characterized as fluid interactions embodied in a highly plastic brain, and that the best model we have for capturing the essence of developmental processes is connectionism. Sections that deal most directly with language development are found in Chapter 3 'Ontogenetic development' and Chapter 7 'Rethinking innateness'. In Chapter 3, two sections related to language are 'The child's sensitivity to speech and language' and 'Learning the past tense'. The first section, 'The child's sensitivity to speech and language' is divided into two subsections, 'Predicting the next sound' and 'Vocabulary development'. In Chapter 7, the section with most relevance to language development, is, 'Twelve arguments about innate representations, with special reference to language'. There are also less developed snippet references to language and language development sprinkled throughout the entire book.

One message of this book that comes through loudly and plainly is that developmental theory needs greater precision. What has been called theory in the past has often been more metaphor than mechanism. The argument is made that connectionism offers us hope of reformation: out with fuzzy, analogical imprecision, and in with connectionist models that are painstakingly honed and refined until they mimic developmental phenomena exactly. In a revealing dialogue quoted at the beginning of the second chapter 'Why connectionism?', a recently converted connectionist (Researcher A) explains the motivation for her/his new affiliation to a colleague (Researcher B).

> Researcher B: ... I thought you believed in constructivism, interactionism, epigenesis and all that *murky* stuff [italics mine].
>
> Researcher A: Oh, I'm still a believer! But the connectionist framework allows me to come up with a much more precise notion of what all that stuff really means. (p. 47)

I think we should feel a common embarrassment over this dialogue, because it fits all too many of us. Driven by hunch and bias, as we so often are, we are led to believing, and in our beliefs we accept models that are based on analogy and metaphor. Researcher A exists in all of us. But why is connectionism our salvation? Consider the overused notion of stage. One does not have to be a neo-Piagetian, constructivist or an interactionist to have evoked the construct of stage to account for developmental phenomena. In addition, we see the construct of stage invoked by such non-interactionists as Radford (1990) and Wexler (1994). *Rethinking innateness* elegantly demonstrates that apparent stages can disguise continuous non-linear change. Chapter 4, 'The shape of change', is by far the best chapter in the book. The beginning of this chapter, containing sections entitled 'Linear change',

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

'Nonlinear monotonic change with linear dynamics', and 'Nonlinear monotonic functions with nonlinear dynamics' should at least be somewhat familiar to developmentalists through non-linear growth curve models (Burchinal & Appelbaum, 1991). Later sections of this chapter might be more familiar to researchers who have had some exposure to dynamical systems approaches to development (Thelen & Smith, 1994). The chapter demonstrates mathematically how continuity can underlie seeming discontinuity in development. The chapter also introduces the reader to concepts in dynamical systems in an extremely well-written and non-intimidating style. This chapter alone is worth the price of the book. It presents a clear challenge to developmental psycholinguists. By showing that even non-monotonic change can be governed by a single equation, the authors argue persuasively that we can no longer accept the convenient ascription of stages. In the future, researchers will be forced to demonstrate that the phenomena they are observing cannot also be accurately captured by non-linear, dynamic alternatives. Even if a researcher rejects connectionism, developmental psycholinguists are going to have to be a great deal more careful in their research. Actually, this chapter underscores a very simple truth. When we deal with problems in defining the shape of change, there is no better remedy than better sampling and more complete data. If we are going to reject a non-linear model, we will have to demonstrate that statistically sufficient amounts of data, adequately sampled, do not fit a non-linear model. Research into the development of regular and irregular morphology has already been spurred by this challenge (Marcus *et al.*, 1992; Kim, Marcus & Pinker, 1994). This field is better for the challenge, and I am sure that more developmental psycholinguists will rise to it in the future. The implications of Chapter 4 go even further. The subsection 'Interacting patterns' shows that research on developmental dissociations in various cognitive domains may be flawed in assuming that these domains are related in a linear, monotonic fashion. Rather, if these domains actually develop in non-linear fashion, apparent dissociation may well disguise an underlying association that changes regularly over time.

Despite my enthusiasm for the successes of *Rethinking innateness*, I see the work as double-edged. Ultimately, I think this book will only serve to worsen polarizations that have been generated and perpetuated in our field. I have two reasons for believing this. The first reason is that the authors obviously intend to make a statement about grammatical development with connectionist models, but have a somewhat unusual perspective on what is significant in the development of grammar. The second reason is that the authors appear to turn a blind eye to well-founded criticisms of connectionist simulations of grammatical development.

Were I reading this book as an outsider to the field of developmental psycholinguistics, I would get the impression that developmental psycho-

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

linguists are interested in the development of the lexicon and past tense verb morphology. Yes, we are interested in those areas, but we are also interested in a great deal more, such as the development of case and agreement systems, the development of negation, WH-questions, subject drop, subject-auxiliary inversion, relative clauses, cleft sentences, constraints on coreference, the semantics of verbs, the argument structure of predicates, phonological processes, metrical structure, and distinctive phonetic features, just to name a few. What developmental psycholinguists see as central to the study of grammatical development is how children develop syntactic structure, morphological form, grammatical features, semantic features, lexical structures, a host of intricate phonological regularities, and how they orchestrate all of these components into one coherent system, employed at incredibly rapid speeds in the production and comprehension systems we see in adults. What the authors find significant with regard to grammatical development, I sometimes find curious. For example, in Chapter 2, 'Why connectionism', the advantages of connectionist representations are discussed:

> 'This spatial framework allows some categories to be distinct and disjoint, but also makes it possible to have representations which lie between category boundaries. An example of this occurs in a phenomenon called sound symbolism... for example words which contain a final *-rl* ('curl', 'unfurl', 'burl', 'whirl', 'twirl' etc.) often evoke the image of circularity... This is the sort of phenomenon which might be exploited by distributed representations'. (p. 95).

I have to say that I find sound symbolism of minimal importance to grammatical development. The syllabic nucleus '-rl' can evoke an image of circularity, but it arguably does not in the words 'earl', 'hurl' and 'girl'. Whereas I have no doubt that sound symbolism exists, it is unclear as to how accounting for it helps us explain much about grammatical development. I should point out that I am truly sympathetic to the notion that submorphemic units play a role in the development of morphology (Rispoli, 1994; in press), but I doubt sound symbolism holds the key.

Whereas sound symbolism seems to hold a significant place for the authors, linguistic theory and what it offers to the study of the development of grammar is denigrated. In a section of Chapter 3 entitled 'Learning the past tense', there is a review of connectionist attempts to model the acquisition of English past tense morphology. In a confused discussion of linguistic theory and research the following remark is made:

> '... much of current linguistic theory in the tradition of transformational generative grammar attempts to postulate highly general rules, such as Move Alpha (*move anything anywhere* [italics mine]), in order to reduce the number of rules needed to characterize so-called Universal Grammar' (pp. 130–1).

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

At this point, the text fails to provide a relevant citation for the concepts of Move Alpha or Universal Grammar. The text badly misrepresents the principle of Move Alpha, which most certainly does not reduce to 'move anything anywhere'. Here, as elsewhere, the anti-linguistics rhetoric is sharp and it is bound to alienate some readers. We must bear in mind that turning up one's nose to research that presents a challenge to one's world view will probably only intensify the worst part of this field's collective behaviour, further contributing to the unfortunate polarization in it. Elevating minutiae such as sound symbolism to an unwarranted status will probably only add fuel to the conflagration.

Throughout the book, simulations of grammatical development are used as illustrations of the important properties of connectionist models. The presentation of simulations of grammatical development is understandably biased in an effort to showcase these simulations in their best light. However, the lack of self-criticism makes further scrutiny by the reader imperative. Connectionist simulations begin with choices about the nature and scope of the input. If these choices are poorly justified, the value of the simulation is greatly diminished. All the simulations of grammatical development presented in this book suffer from this general weakness. Consider the presentation of Elman (1990) in Chapter 2, 'Why connectionism?'. It is claimed by the authors that distributed representations allow for the development of syntactically defined lexical categories:

> 'We see that the network has learned that some inputs have very different distributional characteristics than others, and forms hidden unit representations which places these two groups in different areas of activation space. These groups correspond to what we call nouns and verbs' (p. 95).

The following is an assumption of the Elman (1990) simulation.

> '... from the point of view of the listener, the surface order is the only visible (or audible) part. Whatever the abstract underlying structure be, it is cued by surface forms, and the structure is implicit in them' (Elman, 1990, p. 195).

Whereas it is difficult to argue with the logic of this statement, the simulation does not succeed in demonstrating the power of this logic. The simple sentences that form the input in this simulation contain mutually exclusive sets of nouns and verbs. There was no overlap between these classes. This is a weakness in the demonstration, because, as explained, the emergence of syntactic word classes is solely dependent on co-occurrences with other words in the input strings.

> 'The input representations give no information (such as form class) that could be used for prediction ... Whatever generalizations are true of word

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

> classes must be learned from the co-occurrence statistics…' (Elman, 1990, p. 198).

If the classes of nouns and verbs presented in the input overlapped, for example, by using *smell* as a verb and *smell* as a noun, the usefulness of co-occurrence statistics would either be seriously compromised or more subtly tested. The English language is replete with pairs of nouns and verbs related by zero-derivation, such as *xerox* and *e-mail*. As long as we think of 'nouns' and 'verbs' as LISTS of items, we miss an essential characteristic of syntactic word classes. *Smell* is a noun when it assumes the grammatical features that a noun (phrase) assumes, namely plurality, definiteness and case. *Smell* is a verb when it assumes the grammatical features that a verb assumes, namely TENSE, ASPECT and AGREEMENT. This is old territory for developmental psycholinguists (Maratsos & Chalkley, 1980), and it is unfortunate that a more historical perspective was not taken.

The clear possibility that the results of the Elman (1990) simulation were dependent on assumptions about the input underscores a problem common to Rumelhart & McClelland (1987), Plunkett & Marchman (1993) and Elman (1990). Simulations currently reflect our ideas about language input, and not our knowledge of language input. Take as yet another example the simulation presented in Elman (1993), about which the following is written:

> 'At the conclusion of learning, the network had learned several things: distinctions between grammatical categories: conditions under which number agreement obtained:…how to represent embedded information.' (p. 345).

Although a full list of the items used in Elman (1993) is not given, it appears that a mutually exclusive set of nouns and verbs were once again used, possibly compromising the result of emergent grammatical category distinction. Agreement appears to have been presented only within the context of third person subject sentences, an artificial constraint on sentence input. Moreover, the conclusion is not technically correct. The network did not learn English agreement, because English agreement includes person as well as number. English uses the *-s* suffix on verbs when the subject is singular in the third person, but not first or second person. Person and number are only relevant in the present tense for all verbs but one, *be*, which exceptionally has two different forms in the past. Finally, the input of relative clauses appears always to have contained a relative pronoun. English relative clauses do not always have overt relative pronouns, as in *The boy the girls saw chased a cat the girls didn't see*, and it is impossible to determine what effects relative clauses with non-overt relative pronouns would have had on the results of the simulation.

Connectionist simulations begin with determinations about the nature and scope of the input. If the input to a simulation is preselected to avoid

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

problems, one has instantiated an expert filtering system. Suppose one increased the scope of the Elman (1990) simulation to include pronouns, and therefore sentences like *he chases him*. Because the meanings of words emerge out of co-occurrence statistics in this simulation, *he* and *him* must fall into different classes of words. At the syntactic level *he* and *him* express different grammatical cases, nominative and objective respectively. However, semantically speaking, these are forms of the same word, they mean the same thing, namely a masculine, third person singular referent. In essence, the Elman (1990) simulation avoids the semantic representation of words by attempting to substitute surface structure position for semantics. This theoretical approach resembles that of syntactic bootstrapping in its avoidance of semantic representations and dependence on surface form alone (Rispoli, 1995). However, in order to accomplish the goal of creating word classes by surface structure alone, it appears that the input must be filtered in just the right way. If the simulation's goal is accomplished by avoiding pronouns, then we have the equivalent of a pronoun filter. If this goal is accomplished by avoiding homophones, then we have a homophone filter. If the goal is accomplished by avoiding first and second person referents, then we have a person filter. In fact, the list of what must be avoided is very long indeed.

Consider once again the past tense simulations. All associate past tense forms with the simple, uninflected form of the verb, because the input is structured only to consider these two forms. Of course, there are more forms to the verb paradigm than just these two. All the work of sorting through the verb paradigm and arriving at that one form that means 'past tense' has been done for the simulation. Implied is the fact that the 3rd person singular form does not mean past tense or first person singular, that the progressive form means progressive, and a great deal more. Hidden in these past tense simulations are the constructs of semantics representations, grammatical feature, and paradigm, all of which are developmentally *a priori* to the sorting out of allomorphic variation. What constitutes input in a simulation is essentially external to the simulation and therefore unexplained. The constructs just mentioned, semantics, grammatical feature, and paradigm, are essential to understanding the nature of grammatical development, but, from this book, it appears that the connectionism has nothing new to say about them. Readers of this review might wish to argue that it is unfair criticism to say that researchers who use simulations must be responsible for constructs that are *a priori* to their simulations. After all, one must start somewhere and the middle is as good a place as any to start. However, understanding the nature of the *a priori* will probably have a bearing on the question of what is innate in grammatical development, and language development in general. If one of the purposes of this book is to help us rethink innateness in this domain, it is not unfair to point out that one builds a house from the foundation up.

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

Given the use of these simulations in arguing for the superiority of connectionism, I do not think it unfair to point out that the successes of this book are diminished by the weaknesses of these example simulations. This is especially true when it comes to the area of grammatical development. In their preface, the authors warn us of the following:

> 'We take a connectionist perspective, but we are aware that ours is a specific and significantly enlarged conception of what connectionism is. In some ways, it is our view of what connectionism SHOULD (and hopefully, will) be' (p. xiii).

There is a pledge implicit here. If we put our faith in this vision of connectionism, we will go farther faster in answering our questions. However, if we look at the record as it stands, there is little indication that simulations of grammatical development can make good on this pledge. One might retort that developmental psycholinguistics lacks grand vision, and that developmental psycholinguists love to scrutinize and nitpick. Considering where we have come from, the murky imprecision of the past, it seems inevitable that developmental psycholinguists will want to examine the details of each new simulation. There are other connectionist simulations of grammatical development which are arguably better approximations of reality. Take as an example, Oshima-Takane (in press), which reports a simulation of the acquisition of the first and second person pronoun distinction. This simulation deals honestly with the problem of *a priori* knowledge. It acknowledges that the semantic features of 'self' and 'other', 'speaker' and 'addressee' were *a priori* to, and necessary for, the success of their simulation. The distinction between first and second person pronouns does not arise *de novo*, but is grounded in an already existing semantic foundation. There are also simulations which do not rely on what we THINK input is like. Gobet & Pine (1997) used actual child-directed speech as input to their simulation of the development of early phrasal constituency. The results of their simulation were mixed. There were both failures and successes. But this mixture of results underscores the problems simulations face when real data is used as input. Their demonstration stands on firmer empirical ground than other simulations which do not even attempt to take into account what input is actually like. There are researchers who are working hard to bring connectionist simulations closer to the reality of grammatical development. It is unfortunate that they are overlooked by this book.

In sum, this book succeeds in challenging us on a crucial question: just what constitutes adequate explanation in developmental psycholinguistics? Many of the metaphors of the past are clearly inadequate, and we can no longer afford to rely on them if we are to progress. The book asks us to have faith in connectionism. The book offers tantalizing glimpses of what could

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

be. However, it ignores well-founded criticism off what has been done. The book continues the tradition of polarization through hyperbole that threatens to run developmental psycholinguistics into the ground. For these reasons, the book has the potential for increasing the already substantial divisiveness that mars the field.

REFERENCES

Burchinal, M. & Appelbaum, M. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development* **62**, 23–43.

Elman, J. (1990). Finding structure in time. *Cognitive Science* **14**, 179–211.

—— (1993). Learning and development in neural networks: the importance of starting small. *Cognition* **48**, 71–99.

Gobet, F. & Pine, J. (1997). Modelling the acquisition of syntactic categories. *Proceedings of the 19th Annual Meeting of the Cognitive Science Society*. (265–70). Hillsdale, NJ: Erlbaum.

Kim, J., Marcus, G. & Pinker, S. (1994). Sensitivity of children's inflection to grammatical structure. *Journal of Child Language* **21**, 173–209.

Maratsos, M. & Chalkley, M. (1980). In K. Nelson, (ed.) *Children's language* (Vol. 2, 127–214). New York: Gardner Press.

Marcus, G., Pinker, S., Ullman, M., Hollander, M., Rosen, T. & Xu, Fei. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* **57** (4, Serial No. 228).

Oshima-Takane, Y. (in press). The learning of first and second person pronouns in English. In Jackendoff, R., Wynn, K. & Bloom, P. *Studies in memory of John Macnamara*. Cambridge, MA: MIT Press.

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

Radford, A. (1990). *Syntactic theory and the acquisition of English syntax*. Cambridge, MA: Basil Blackwell.

Rispoli, M. (1994). Pronoun case overextensions and paradigm building. *Journal of Child Language* **21**, 157–72.

—— (1995). Missing arguments and the acquisition of predicate meanings. In M. Tomasello & W. Merriman (eds) *Beyond names for things: young children's acquisition of verbs*. (331–52). Hillsdale, NJ: Erlbaum.

—— (in press). Pronoun case errors: me or my: two different patterns of pronoun case errors. *Journal of Speech and Hearing Research*.

Rumelhart, D. & McClelland, J. (1987). Learning the past tenses of English verbs: implicit rules or parallel distributed processing? In B. MacWhinney (ed.) *Mechanisms of language acquisition* (195–248).

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

Wexler, K. (1994). Optional infinitives, head movement and the economy of derivations. In D. Lightfoot & N. Hornstein (eds) *Verb movement*. Cambridge, England: C.U.P.

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

Connectionism, theories of learning, and syntax acquisition: where do we stand?

NICK CHATER
University of Warwick

AND

MARTIN REDINGTON
University College, London

This commentary concerns two issues: the contribution of connectionism and theories of learning to developmental psychology; and the status of connectionist models of syntax acquisition.

Connectionism and theories of learning

'from this book, it appears that connectionism has nothing new to say about [grammatical development]' (Rispoli, p. 12).

For developmental psychology, and in particular grammatical development, what is new about connectionism? One answer, advanced in RI, is that connectionism represents a model of computation using idealized neural hardware. A complementary answer, from an information processing perspective, is that connectionism illustrates the power of computational learning methods. Over the last forty years, there have been enormous theoretical and practical advances in computational accounts of learning, which developmental psychology has largely ignored. Connectionism, as persuasively applied in RI, makes some of this work accessible to developmentalists. We hope that RI will also alert developmental psychologists to other areas of progress in the formal study of learning.

A major theoretical step in understanding how learning is possible was made with the development of Kolmogorov complexity theory (e.g. Solomonoff, 1964*a*, *b*). The idea is that learning can proceed by searching for simplicity (rather as has been proposed in perceptual organization, Leeuwenberg, 1971; Leeuwenberg & Boselie, 1988). The simplest 'explanation' for a set of data (e.g. corresponding to sensory and motor information) is guaranteed both to provide reliable predictions, and to be the most likely to be true. Learning mechanisms are viewed as searching for the simplest explanation; simplicity provides a standard against which different grammars, hypotheses, or theories can be assessed. Moreover, the simplicity of an explanation can be measured straightforwardly in terms of the code length required to specify that explanation. Theories of learning based on

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

Kolmogorov complexity have led to important developments in statistics and machine learning as well as having explicit (e.g. Zemel, 1993) or implicit (Chater, 1995) links with connectionist learning as discussed in RI. Thus, specific connectionist models can be viewed as examples of the general theoretical claim: that learning operates by searching for simplicity.

It turns out that different simple explanations must generally agree on their predictions (Li & Vitanyi, 1997). This means that if different language learners attain different grammars, which capture the language well (and hence provide simple explanations of it), then their grammatical intuitions will typically agree. Thus, to explain why the final state of human language acquisition shows common intuitions across individuals, one need not assume that the goal of language acquisition is attaining a 'correct' grammar; it is enough that learners attain grammars which provide simple explanations of linguistic data – even though the specifics of these grammars may vary widely across speakers.

This casts apparently negative results from another branch of the mathematical study of learning, formal learning theory (Gold, 1967; Osherson, Stob & Weinstein, 1985), in a new light. These results concern the difficulty of converging, in the limit, on a single 'correct' grammar, particularly given only positive evidence. But this may not be the problem confronting the child.

Moreover, along similar lines, it is often assumed language learning will be rendered impossible by the problem of over-general grammars (such as, 'any word can follow any other'). Such grammars are trivially consistent with the linguistic data, and apparently can only be ruled out by negative evidence, the role of which is empirically controversial. Thus, it is concluded that there must be very strong innate constraints on language acquisition (e.g. Baker & McCarthy, 1981; Chomsky, 1980; Pinker, 1979, 1984, 1989). Learning by simplicity, however, rules out overgeneral grammars immediately, because, while consistent with the linguistic data, they do not explain that data simply. Simple explanations must exploit the structure in the data; and overgeneral explanations are oblivious to that structure. Thus, if learning is a search for simplicity, overgeneral grammars are disfavoured, even given only positive evidence.

Overall, viewing language acquisition as a search for simplicity changes the learning problem into a more tractable form. Of course, innate knowledge may, nonetheless, play an important role in language acquisition. But by studying the potential of connectionist and other formal frameworks for studying learning, we can investigate how innate knowledge and learning mechanisms might jointly contribute and interact in language acquisition.

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

Connectionism and syntax acquisition

Elman's (1990) connectionist simulations, discussed in RI, and by Rispoli, are based on what Elman calls 'simple recurrent networks' (SRNs). SRNs are typically given a sequence of inputs (a 'word' at a time), and the task is to predict the next input. The degree to which the predictions respect the grammar underlying the inputs measures how much linguistic structure that the network has learned.

Rispoli is concerned that SRNs (and the other models in RI) avoid the hard problems in syntax acquisition, primarily because they use simplified artificial input, not natural language. Specifically, in word class acquisition, the problem is drastically simplified by using small vocabularies in which each item has a single syntactic category, whereas in natural language, many word forms have several syntactic categories.

While Rispoli's concerns are appropriate, Elman's simulations, like the other simulations in RI, illustrate that connectionist mechanisms are capable of learning linguistically relevant structure, from language-like input. Naturally, the closer the approximation of the input to that received by the child, the stronger the conclusions that can be drawn. But drawing negative conclusions on the basis of over-simplified input is unjustified.

Indeed, work with (transcribed) spoken language, shows that distributional information provides a powerful cue to word class membership. Redington, Finch & Chater (in press) clustered words based on the statistical similarity of their contexts (a 'context' is just of the next and previous two words). Using adult language from the CHILDES corpus (MacWhinney & Snow, 1985), they found close agreement between the classification derived using this method (shown in Figure 1) with the most frequent syntactic category associated with each word token, according to a standard grammatical analysis. Figure 2 shows finer level structure, illustrating that distributional information is informative about lexical semantics (see also Landauer & Dumais, 1997; Lund & Burgess, 1996). Moreover, these techniques have been shown to work in other languages, including Chinese (Redington, Chater, Huang, Chang, Finch & Chen, 1995) and (in our unpublished work) German.

It is unclear how successfully distributional methods can extract multiple syntactic categories for syntactically ambiguous word forms, a problem that Rispoli highlights. Neither do distributional methods perfectly classify words by most common syntactic class. However, they illustrate that the distributional information present in natural language, when considered in isolation, conveys much more information about word class membership than intuition would suggest.

More generally, Rispoli's concerns raise the question: do SRNs provide a potential model for human language acquisition? On the positive side, SRNs

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

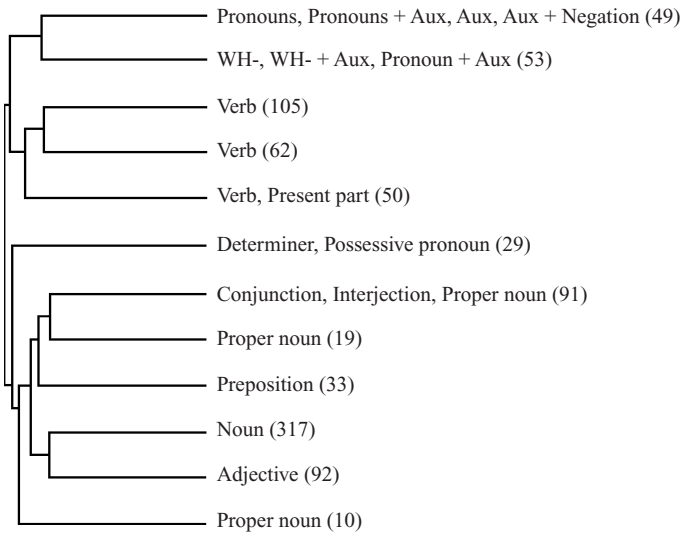


Fig. 1. The high level structure of a cluster analysis of words based on their distributional similarity in the CHILDES corpus. Specifically, the cluster diagramme has been 'cut' into discrete categories (at a similarity level of 0·8, see Redington, Chater & Finch, in press, for details). The categories have then been labelled by hand with the syntactic categories to which they best correspond. These labels are highly reliable, in the sense that almost all items in a particular cluster have the relevant syntactic category as their most frequent reading. (Only clusters which include ten or more words are included here.)

can pick up limited recursive structure, when trained on simple artificial languages (Christiansen & Chater, 1994, in press; Elman, 1990, 1991, 1993). Indeed, with centre-embedded and cross-dependency constructions, Christiansen & Chater (in press) found that SRN performance breaks down at the same depth of embedding as human performance (Bach, Brown & Marslen-Wilson, 1986). On the negative side, SRN performance declines steeply as vocabulary size increases. This is because the method learns by prediction, which is increasingly difficult when vocabulary size is large. This suggests that, to the extent that SRNs may be useful models of language acquisition, they must operate over preclassified input – i.e. a sequence of syntactic classes, rather than individual lexical items. Applying SRNs on real-world input represented as a stream of syntactic classes is an interesting project for future research. But, of course, language acquisition involves not merely learning recursive structure, but a set of immensely intricate linguistic constraints – a formidable task.

So where do connectionist models of syntax acquisition stand? SRNs represent an interesting direction for learning recursive structure; and

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

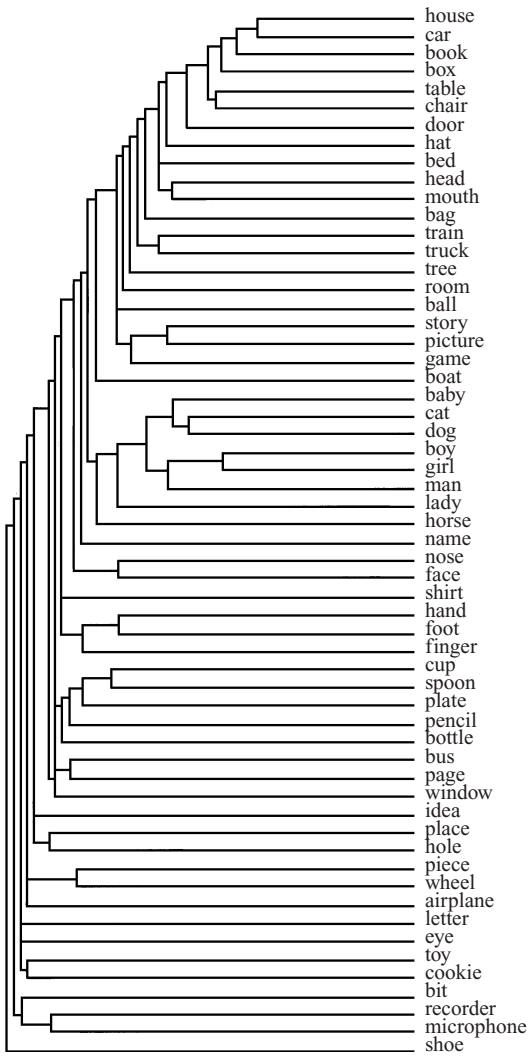


Fig. 2. A subcluster of nouns from Fig. 1. Notice that these show a considerable degree of fine-grained semantic clustering. For example, 'table' is clustered with 'chair', 'hand' with 'foot', and so on.

distributional methods can categorize word tokens into syntactic categories. But these computational techniques are currently extremely limited compared to human language acquisition. Further research is required to clarify where the difference lies – because the child has better learning techniques,

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

exploits additional sources of information (e.g. phonological, prosodic, semantic and pragmatic), possesses deep innate constraints on which grammars can be entertained, or any combination of these. By focusing attention on developments in connectionist research on language acquisition, it may be hoped that both RI, and Rispoli's stimulating commentary, may encourage such research.

REFERENCES

Bach, E., Brown, C. & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: a psycholinguistic study. *Language and Cognitive Processes* **1**, 249–62.

Baker, C. L. & McCarthy, J. J. (1981) (eds). *The logical problem of language acquisition*. Cambridge, MA: MIT Press.

Bullinaria, A. & Cairns, P. (eds). (in press). *Connectionist models of memory and language*. London: UCL Press.

Chater, N. (1995). Neural networks: the new statistical models of mind. In J. P. Levy, D. Bairaktaris, J.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review* **103**, 566–91.

—— (1997). Simplicity and the mind. *The Psychologist*, November, 1997, 495–98.

—— (1998). The search for simplicity: a fundamental cognitive principle? *Quarterly Journal of Experimental Psychology*, in press.

Chater, N. & Conkey, P. (1992) Finding linguistic structure with recurrent neural networks. In *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society* (402–7). Hillsdale, NJ: Erlbaum.

Chomsky, N. (1980). *Rules and representations*. Oxford: Blackwell.

Christiansen, M. & Chater, N. (1994). Generalization and connectionist language learning. *Mind and Language* **9**, 273–87.

—— (in press). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* **00**, 000–000.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science* **14**, 179–211.

—— (1991). Distributed representation, simple recurrent networks, and grammatical structure. *Machine Learning* **7**, 195–225.

—— (1993). Learning and development in neural networks: the importance of starting small. *Cognition* **48**, 71–99.

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

Gold, E. M. (1967). Language identification in the limit. *Information and Control* **10**, 447–474.

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* **104**, 211–40.

Leeuwenberg, E. (1971). A perceptual coding language for perceptual and auditory patterns. *American Journal of Psychology* **84**, 307–49.

Leeuwenberg, E. & Boselie, F. (1988). Against the likelihood principle in visual form perception. *Psychological Review* **95**, 485–91.

Li, M. & Vitanyi, P. (1997). *An introduction to Kolmogorov complexity and its applications* (2nd Edition). Berlin: Springer Verlag.

Lund, K. & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers* **28**, 203–8.

MacWhinney, B. & Snow, C. (1985). The child language data exchange system. *Journal of Child Language* **12**, 271–95.

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

Osherson, D. N., Stob, M. & Weinstein, S. (1985). *Systems that learn*. Cambridge, MA: MIT Press.
Pinker, S. (1979). Formal models of language acquisition. *Cognition* **7**, 217–83.
—— (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
—— (1989). *Learnability and cognition : the acquisition of argument structure*. Cambridge, MA: MIT Press.
Plunkett, K. & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: implications for child language acquisition. *Cognition* **38**, 43–102.
—— (1993). From rote learning to system building. *Cognition* **48**, 21–69.
Redington, M. & Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends in Cognitive Sciences* **1**, 273–81.
—— (in press). Connectionist and statistical approaches to language acquisition: a distributional perspective. *Language and Cognitive Processes*.
Redington, M., Finch, S. & Chater, N. (in press). Distributional information: a powerful cue for acquiring syntactic categories. *Cognitive Science*.
Redington, M., Chater, N., Huang, C., Chang, L.-P., Finch, S. & Chen, K. (1995). The universality of simple distributional methods: identifying syntactic categories in Chinese. In *Proceedings of the Cognitive Science of Natural Language Processing*, Dublin, 1995.
Solomonoff, R. J. (1964*a*). A formal theory of inductive inference, part 1. *Information and Control* **11**, 1–22.
—— (1964*b*). A formal theory of inductive inference, part 2. *Information and Control* **11**, 224–54.
Zemel, R. S. (1993). *A minimum description length framework for unsupervised learning*. Unpublished doctoral dissertation, Department of Computer Science, University of Toronto, Canada.

The windmills of your mind: commentary inspired by Cervantes (1615) on Rispoli's review of *Rethinking innateness*

NICK C. ELLIS
University of Wales, Bangor

Rethinking innateness (RI) is a quixotic book, enthusiastic and visionary. It is hard not to be provoked as each author races full tilt towards their chosen combat. Some readers, like Fodor (1997), respond in kind.[1] Rispoli is more measured. Nonetheless, every reader chooses their own windmills to elevate into giants, and each believes their battle righteous. Are Elman *et al.* 'building castles in the air', or is Rispoli 'taking the wrong sow by the ear'?

Rispoli identifies the major virtues of RI as its contributions towards,

[1] 'Fly not, cowards and vile beings, for a single knight attacks you…Though ye flourish more arms than the giant Briareus, ye have to reckon with me.'

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

variously, the precision of developmental hypotheses, the sophistication of theories of non-linear development and the realization of connectionism as techniques for formally testing theories as mechanisms rather than metaphors. Agreed, but to my mind he underplays cognitive neuroscience's emphasis on the biology of the development of mind: 'knowledge' must be understood in explicit, systematic terms that can be implemented in some kind of biologically plausible machine – the overview in Chapter 5 of brain development and consequent dismissal of representational innatism is particularly compelling. He ignores the extreme interdisciplinarity of a book jointly authored by six researchers from quite different specialisms. Thus he sells short emergentism as the alternative framework for cognitive development: interactions occurring at all levels, from genes to environment, give rise to emergent forms and behaviour; dramatic effects can be produced by small changes in timing of events; these complex interactions can be explored through dynamic models using a synthesis of connectionist modelling and developmental neurobiology.

What of the failings Rispoli identifies? He correctly describes the linguistic content of RI as both scant (what critics of mine have referred to as 'just that lexical stuff') and caricatured. His sentence about the 'great deal more' that concerns developmental psycholinguistics is almost as quotable as Fodor's 'no examples are given of how, even in sketch, an attested linguistic universal might be explained in this way' (Fodor, 1997, p. 4). The connectionist simulations of language featured in RI are almost exclusively those undertaken by Elman and Plunkett. It is unfortunate that they 'ventured all their eggs in one basket' in this fashion. Perhaps they believed that these models served to illustrate the advantages of connectionism, as indeed they do. But RI is now being reviewed, and found wanting, as the definitive statement on connectionist approaches to language development, rather than as a book, purposefully entitled *Rethinking innateness*, which argues an emergentist developmental perspective while proffering some examples from the domain of language.

Rispoli's chief criticism is as follows: 'Connectionist simulations begin with choices about the nature and scope of input. If these choices are poorly justified, the value of the simulation is greatly diminished. All the simulations of grammatical development presented in the book suffer from this general weakness'. These important observations echo (i) Lachter & Bever (1988) – does the input data 'cryptoembody' rules as TRICS (The Representations It Crucially Supposes) and (ii) Carroll (1995) – do the modeller's combined choices of problem subspace, architecture and input effectively solve the problem before learning has even begun? Rispoli illustrates this by scrutinising Elman's (1990) deriving of syntactic class clusters from sequential word dependencies: this simulation used simplified sentences from a toy English where nouns and verbs constituted mutually exclusive sets; but, since noun

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

and verb classes overlap in real English, it is a 'clear possibility that the results of the Elman (1990) simulation were dependent upon assumptions about the input'.

It is a possibility. But subsequent simulations have shown him wrong in claiming that 'to accomplish the goal of creating word classes by surface structure alone, it appears the input must be filtered in just the right way'. Redington & Chater (1997) report analyses of very real English: (i) 40 million words from USENET newsgroups, replete with 'typographical errors, ungrammatical sentences,and all manner of idiosyncratic stylistic quirks' and (ii) 2·5 million words of child-directed English speech from CHILDES, again very noisy and unfiltered, where *mommy* and *mummy* are effectively different words and words like *poo* and *smell* regularly feature as both nouns and verbs. Far from a 'wild-goose chase', these computational analyses of natural language show word distributions to be highly informative about syntactic category. This does not deny important questions of how children learn that some lexical items have multiple syntactic class membership, or that homonyms have multiple meanings. But these are exactly the issues that are best addressed by looking to the data.

Rispoli is misleading when he claims 'simulations currently reflect our ideas about language input, and not our knowledge of language input'. Emergentists use connectionist and statistical techniques to determine what patterns are latent in the input. Their acknowledgement of the importance of input engenders an intimate concern for their models' exposure to representative input. Indeed, connectionists and corpus linguists work in close collaboration, sometimes to the degree that, like MacWhinney (1995, 1998), they are apparently the same person[2]! It is a shame that the restricted input data of the particular simulations used in RI led Rispoli so to misapprehend the patterns and agreements that underpin this research. Getting to know emergentism, like getting to know language, like cognition in any complex domain, requires large amounts of experience. The following reviews are good leads: Redington & Chater (1997), Seidenberg (1997), Plunkett (1998), Ellis (in press), MacWhinney (in press).

Rispoli rightly complains about the constrained input in Elman (1993). True, this did not learn all the intricacies of English agreement; it was intended as a demonstration, focused on 'starting small', not on the totality of English grammar. Nevertheless, Rispoli is entirely wrong in claiming 'it is impossible to determine what effects relative clauses with non-overt relative pronouns would have had on the results of the simulation'. Perhaps for the casual reader, but such questions are exactly the stuff of research which runs simulations with different data experiences to determine the emergent consequences.

[2] 'He has an oar in every man's boat, and a finger in every pie.'

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

When next addressing semantics, Rispoli demands that sequence analysis 'eat its cake and have its cake' – pronouns like *he* and *him* should fall into different nominative and accusative syntactic cases whilst simultaneously belonging to the same masculine, third person, singular, semantic class. It is wrong to presume that we 'may as well expect pears from an elm'. Recent distributional analyses of English (Landauer & Dumais, 1997) derive lexical semantics effective enough to pass university-level tests of English competence from 'simple bags of words' – unordered frequency profiles of co-occurring words in 30000 text samples roughly 150 words long. Contra Rispoli, no long list of what must be avoided is necessary. Instead, syntactic and semantic information result from different grains of analysis of the same language input – local ordered word co-occurrence data produces syntactic information, more global word co-occurrence informs semantics. For lexical syntax and semantics alike, 'tell me thy company, and I will tell thee what thou art'.

That semantics emerges from lexical co-occurrences in no way contradicts that semantics is ultimately a matter of grounding. There is redundancy. It is the reflections of the world in language that makes language learnable. Cognitive linguistics encourages emergentism/connectionism to turn to simulating the co-development of representations of language, perception, attention and motor control. Progress here will be necessarily slower since it rests on a proper understanding of human cognition.

I have argued with Rispoli, and I have agreed with him. Enough. What might Cervantes now advise? Perhaps simply: 'patience, and shuffle the cards'. Vale.

REFERENCES

Carroll, S. E. (1995). The hidden danger of computer modeling: remarks on Sokolik and Smith's connectionist learning model of French gender. *Second Language Research* **11**, 193–205.

de Cervantes, S. M. (1615). *Don Quixote*. Translated by John Ormsby 1829–1895. Available gopher://gopher.vt.edu:10010/02/62/1.

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning* **50**, 000–000.

Fodor, J. A. (1997). Review of the book Rethinking innateness: A connectionist perspective on development. *Times Literary Supplement*, May 16, 1997.

Lachter, J. & Bever, T. (1988). The relation between linguistic structure and associative theories of language learning: a constructive critique of some connectionist learning models. *Cognition* **28**, 195–247.

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* **104**, 211–40.

MacWhinney, B. (1995). *The CHILDES project: tools for analyzing talk*. Second edition. Hillsdale, NJ: Erlbaum.

—— (ed.) (1998). *The emergence of language*. Hillsdale, NJ: Erlbaum.

Plunkett, K. (ed.) (1998) Connectionist models of language [special issue]. *Language and Cognitive Processes* **13**, 000–000.

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

Redington, M. & Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends in Cognitive Sciences* **1**, 273–81.
Seidenberg, M. S. (1997). Language acquisition and use: learning and applying probabilistic constraints. *Science* **275**, 1599–1603.

Computational models and *Rethinking innateness*

JEROME A. FELDMAN
ICSI and UC Berkeley

On the whole, I quite like both the target book (RI) and the review by Matthew Rispoli, but there are some additional observations that might be of interest to the readers of this journal. The whole concept of 'truth through disputation' is alien to my scientific tradition and I agree with Rispoli that RI is not helped by the polemical tone.

There is a companion volume and software suite by Kim Plunkett & Jeff Elman (1997), *Exercises in rethinking innateness*, which we have used in an undergraduate cognitive science course. This is, in my opinion, the best source for understanding the main point of RI, which I summarized for the class as: the book tries to show that PDP learning techniques have advanced to the point where we need not assume that *tabula rasa* learning of language must be ruled out. The *Exercises in RI* book and particularly doing the on-line examples give the students direct intuition about this claim. Our students were impressed by the Tlearn system, but understood its limitations and were not convinced either of the main claim of RI or of the radical nativist alternative.

Rispoli joins the RI authors in identifying connectionist models with PDP networks. Much fine work continues to be done in this tradition, but it is not the only game in town. There is a continuing body of work that also uses connectionist modelling techniques but employs explicit structures based on what is known about the neural structure in the brain[1]. This 'structured' or 'localist' approach to connectionist modelling is neutral with respect to nature/nurture controversies – any postulated initial structure can be (and is) built into the model. This approach is widely used in modelling language use and acquisition and also captures well the spreading activation results in psycholinguistics. For example, Regier has shown how a simple model of the visual system can support the learning of spatial relation terms from a wide range of languages. A recent dissertation by David Bailey (1998) used similar techniques to model (again cross-linguistically) the acquisition of verbs of simple hand motion and the accompanying conceptual structure. A companion thesis by Srinivas Narayanan (1998) suggests how the same compu-

[1] Our work in this area can be accessed through the web site: www.icsi.berkeley.edu/NTL/

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

tational primitives can serve as a foundation for metaphor and for linguistic aspect. Probably the best introduction for readers of this journal is the book by Terry Regier (1995), *The human semantic potential*, which is part of the same Elman series from MIT press. The most important point is that we are not reduced to a dialectic between extreme empiricist and nativist theories of language acquisition.

REFERENCES

Bailey, D. (1998). *When push comes to shove : a computational model of the role of motor control in the acquisition of action verbs*. Unpublished Ph.D. dissertation, UC Berkeley, also TR 97-041, International Computer Science Institute.

Narayanan, S. (1998). *KARMA : knowledge-based action representations for metaphor and aspect*. Unpublished Ph.D. dissertation, UC Berkeley.

Plunkett, K. & Elman, J. (1997). *Exercises in rethinking innateness*. Cambridge: MIT Press.

Regier, T. (1995). *The human semantic potential*. Cambridge: MIT Press.

Commentary on Rispoli's review

BRIAN MACWHINNEY

Carnegie Mellon University

The debate between nativism and empiricism has been the central issue in developmental psychology for well over three centuries. Nowhere has the opposition between nature and nurture been felt more keenly than in the study of child language development. At times the debate has proven useful, or at least entertaining. But there are those, like Matthew Rispoli, who worry that a fixation on this issue may reinforce a polarization that will eventually 'run developmental psycholinguistics into the ground'.

Rispoli chastises the authors of *Rethinking innateness* for violations of academic etiquette in this ongoing debate. Reasonably enough, Rispoli holds that intellectual argumentation should be governed by two basic principles. First, one should not overstate one's position. Secondly, one should thoroughly understand and correctly represent any position that one criticizes. Rispoli faults the connectionists on both counts. However, Rispoli fails to add the important observation that the violations of academic etiquette did not actually begin in the connectionist camp. For many years, the formalists have been able to avoid misrepresenting the positions of the connectionists and the functionalists by simply ignoring their existence. Given this, I would add a third basic rule of academic etiquette: one should acknowledge the existence of opposing views. On balance, it would seem to me that violations

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

of academic etiquette are well distributed across all parties to the debate between nativism and empiricism.

Human nature being what it is, I doubt that progress toward a resolution of these issues can be achieved by an attempt to enforce good academic etiquette. Instead of trying to influence our colleagues through overt correction (Gold, 1967), I recommend that we simply change the subject of the dialogue. I would say that the true contribution of *Rethinking innateness* is not its stalwart defense of connectionism or its exploration of biological processes, but rather the way in which it opens up dialogue on a topic that can serve as a fundamental replacement for the worn-out debate between nativism and empiricism. This is the notion of emergence. As we learn in Chapter 1, the genes do their work not through direct determination of specific neural or behavioural structures, but through indirect and emergent determination. There is no gene that tells the honey bee how to build hexagonal cells in the honeycomb. The bee simply deposits small drops of honey encased in wax and the operation of packing constraints gives rise to the hexagonal pattern of the honeycomb.

What is true for a simple organism like the honey bee is even more true for complex organisms like humans. The development of the columnar organization of visual cortex in man is not the result of some direct genetic determination, but an emergent property of the competition for connections between afferent processes from the two visual fields. If old, stable structures such as the visual system are determined by emergent forces, why would we want to believe that the new system of abilities that underlie human language should be any less subject to plastic expression?

In fact, emergentist accounts of language acquisition have now been offered in many areas. Consider the papers from a recent conference (MacWhinney, 1998) in which we find emergentist treatments of phonology (Stemberger & Bernhardt; Gupta & Dell; Plaut & Kello; Aslin, Saffran & Newport), word learning (Merriman; Golinkoff, Hirsh-Pasek & Hollich; Smith), social influences (Snow), sentence comprehension (MacWhinney; MacDonald,; Allen & Seidenberg; Miikkulainen & Mayberry), and grammatical structure (Goldberg; Bates & Goodman; Givón and Elman). Some of these treatments rely on connectionist models, but there are other emergentist frameworks as well, including Optimality Theory, dynamic systems theory, competition, functionalist linguistics, interactional theory, and embodied cognition.

What is remarkable about the emergentist approach is the fact that it extends equally well to many aspects of formal theory. By emphasizing the way in which a small set of universal parameters can give rise to a diverse set of phenotypic language types, Chomsky has advanced a fundamentally emergentist argument. The minimalist programme further develops this approach by arguing that, from a minimal set of basic principles, a great

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

variety of complex linguistic patterns can be derived. If even Chomskyan analysis is compatible with emergentism, what views are not? The answer is that emergentism opposes itself most clearly to analysis and modelling through stipulation. In old-style Artificial Intelligence (AI) one could wire-up a production system to model virtually any human behaviour. The system would be brittle and full of *ad hoc* assumptions, but it would generate the required behaviour. This is exactly the type of thinking that generativists have been trying to escape for nearly 30 years. By rejecting unrestricted rewrite rules and by moving continually to constrain the shape of grammar, generativists have been waging their own war against mere stipulation.

Given their mutual interest in emergentism, it is remarkable that the gap between generativists and connectionists remains so unbreachable. As I see it, the only way to mend this rift is to develop concrete mechanistic accounts that show how formal complexity emerges from basic biological principles that control the shape of human cognition. The first few attempts to demonstrate these linkages will run into pretty rough sledding. For example, attempts to derive the effects predicted by levels phonology (*mouses-hater) from the development of lexical patterns have generated controversy without consensus. But the tenor of these arguments is non-polemic and scientifically productive.

Rethinking innateness has succeeded in putting a spotlight on these difficult problems. Although it does little to cool the overheated rhetoric of the nature–nurture debate, I believe that the ideas it developed had to appear in a coherent way in print to allow us to move on with the work of constructing a more comprehensive emergentist approach.

REFERENCES

Gold, E. (1967). Language identification in the limit. *Information and Control* **10**, 447–74.
MacWhinney, B. (ed.). (1998). The emergence of language. Mahwah, NJ: Erlbaum.

A sunny view of polarization

MICHAEL MARATSOS
University of Minnesota

There are a number of ways to react to the whole problem of polarization that Rispoli's review brings up. As a first reaction, like Rispoli, personally, I deplore the polarization. I find neither side likely to be right enough about everything to be entitled to require the kind of all-or-nothing allegiance each

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

side seems to demand. My complaints about the two sides are more basic than Rispoli's rather specific criticisms of Elman *et al*. Unfortunately, going on to outline what seems generally useful and less useful in each approach in a short space like this basically consists of saying 'I like this' or claiming 'this is obviously wrong' with a little elaboration. This does not seem useful for such complicated problems, so I will leave my general complaint about the all-or-none polarization simply stated outright.

But life is complicated. Accompanying these anti-polarization sentiments are reasons for simultaneously being, in part, more sympathetic to polarization. Some of the polarization is intellectually justified. The central propositions of much current connectionist work, and of much current Chomskyan work are, quite simply, non-negotiable and incommensurable. It is wrong to think they could work out a compromise on all these matters. For example, Elman *et al*. practice what I will call, non-pejoratively, reductionist connectionism. This view holds (for example) that so-called symbols like noun, verb, or noun-phrase, or sentence are just epiphenomenal misunderstandings we have, of learned network arrangements of non-linguistic primitive elements. The linguistic symbols are no more real to reductionist connectionists than are 'objects' to physicists who think of 'objects' as our limited perceptual system's misunderstanding of underlying electro-magnetic energy configurations. Chomskyans (among others), in contrast, believe that 'noun' and 'verb' and determiner phrase and various sentence constituents, are not only real mental primitives, but are innately given inherently linguistic primitives of the mind. The difference between these views is approximately as great as the difference between views that humans were created by divine planned agency versus views that humans evolved by chance and natural selection from single-celled organisms. (Though the differences about grammatical acquisition probably fail to have quite the social and cosmological consequences of the latter dispute.) These are non-negotiable differences, and the polarization here is real.

Another basis for a partly sympathetic view of the polarization comes from reading sources in the history of science like Feyerabend (1975). These sources give me, at least, considerable respect for Feyerabend's conclusion that in science 'anything goes' (whether we personally like it or not). Scientific history is full of cases of people being unreasonable and good things coming out of it. Our official doctrine that reason *per se* should be the primary arbiter of our conduct does poorly against the fact that acquiring higher degrees does not mean people stop being human beings. This in turn means that sometimes conduct that is non-ideal by many standards, turns out to work pretty well.

In the case of grammar theory polarization, one resulting optimistic view works this way: in reality, the problems of grammatical development are very difficult. Most or all of the relevant central representations and mechanisms

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

involved are highly removed from direct behavioural expression. As a result, it is very difficult for empirical data to address basic questions in any simple and direct way. This can be highly dispiriting, and make it difficult to continue doing work.

Polarization into social groups, conversely, gives people heart. They can believe their group is morally right in the beliefs it holds, and in the way it proceeds to investigate questions, and this belief can be strengthened greatly by the fact that all the people they talk to feel the same way. Having a common enemy further energizes and unites people.

Furthermore, as modern cognitive psychological work in decision-making shows, people are not really very good at dealing with complicated and mixed situations (which I think grammar and grammatical development obviously are). It is cognitively difficult to deal with complicated problems and situations, because of the limited processing space available in the bottleneck of conscious thought. People do all sorts of things to simplify the matter as a result. For example, decision-making studies show that people like to simplify things initially by applying some single initial uniform criterion by which a great many alternatives are immediately eliminated. They do not exhaustively inspect the total of good and bad points of each view in this initial sweep. Rather, they carry out more careful evaluation of the remaining alternatives only after making the initial eliminations on the first, relatively simple basis. This initial pruning procedure becomes even more likely as the problems and alternatives become more complicated. Obviously this is what we see in the arguments of our field. One eliminates a wide range of opponent views because of one or two serious problems they have, which saves one from having to look at anything good they might also have. One then settles down to look at good and bad points only within the remaining domain of alternatives (essentially one's own school).

Another simplifying device is representativeness. People naturally tend to assume that a cognitive domain is homogeneous, and that any part of it represents the whole. Thus, one can concentrate on a few findings or points that support one's unified view, and take it that the whole problem is thereby representatively characterized. This makes it easier to contemplate the problem, and also increases the possible significance of a single finding one might obtain.

So polarization of the type commonly found in our field – and others – gives people heart and simplifies matters in ways congenial to their thought. Of course, this is not a good thing if there is NO merit at all in the view being pursued. Nazis had plenty of heart, energy and unity of view, but we do not find this endearing. Less seriously (more on our scale, in other words), behaviourists also had plenty of spirit, energy, conviction and unity and attracted many bright persons to their side. But it cannot be said, currently, that this worked out all that well.

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

Furthermore, the polarized approach typically requires that one's side win total victory. Often in the history of psychology, precisely because people are complicated, no total view wins complete victory. Then everyone decides nothing was accomplished because nobody won the contest cleanly, as was required by the polarized approach.

In fact, however, the optimistic view of our situation works this way: neither major side is totally right, and thus by the highest intellectual-ethical standards we tend to bring out, neither is entitled to the kind of polarized intellectual and methodological views they would urge (force?) upon everyone else. But on the other hand, it seems likely that each side does have a good deal to work on that is either right or at least promising. If this is so, in the long run, it is better that people find a social situation that allows them to pursue the promising aspects of their views. People being what they are, this may require them to be unreasonable. Probably our public ideal is that people should be objective judges, and act accordingly. In reality, what people typically are is lawyers who want to argue for their own side and ignore the good points of the other side. What makes this work as well as it does is freedom of speech, the freedom of opposing points of view to produce new arguments and findings with which, in the long run, something useful may be partly sorted out. This makes life less than ideal for many of us who find the resulting social-intellectual situation unpleasant. It is not GUARANTEED to produce good results. But sometimes it does, and with luck this will be one of those times.

What do connectionist simulations tell us?

FERNAND GOBET
University of Nottingham

In his review, Rispoli's main concern is that Elman *et al.*'s book will aggravate the degree of polarization in developmental psycholinguistics. I cannot really comment on this worry, as developmental psycholinguistics is not my field. Instead, I will discuss some questions more related to my background – the role of computational modelling in Elman *et al.*'s approach.

Elman *et al.*'s ambitious goal is to propose theories of cognitive development that are grounded in our knowledge of biology. This is of course what the great Jean Piaget tried to achieve during his lifetime – unsuccessfully, as we know. Elman *et al.*'s advantage over Piaget is that they have a set of computational tools, connectionnism, which both allows them to specify theories precisely and to study complex behaviours (such as epigenesis,

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

where innate and environmental factors interact to create new levels of complexity) that are just beyond the (unaided) human mind. Even though I will highlight some of the weaknesses of their approach below, I should emphasize that reading their book was an exciting and enjoyable experience.

As noted by Rispoli, there are important problems with the simulations reported by Elman *et al*. Rispoli focuses on simulations of past tense acquisition and syntax acquisition, but the problems are by no means limited to these areas. I will briefly consider two recent developments in neural net research, one taken from the field of language acquisition, and one from elsewhere, which underscore some of the difficulties of the simulations discussed in the book.

The influential simulations reported first in Elman (1993), stressing the 'importance of starting small', play a significant role in the book, because they support one of its key ideas, that time-related constraints on perceptual capacity and memory capacity may facilitate the acquisition of language (or of other complex skills) in that they limit the search space faced by biological brains and by artificial neural networks. Elman found that a recurrent network could not acquire a simple, artificial grammar containing relative clauses when full sentences were given as input, but that it could do so either when the complexity of the input was increased gradually, short sentences being presented first, or when the temporal window with which the network could process information was limited to three or four words at the beginning and then progressively extended to its full size. However, Rohde & Plaut (1997), using essentially the same type of recurrent network as Elman, were unable to replicate this finding. They found that their network could learn better when it was presented with full sentences than when presented with an incremental regime. They concluded that, if anything, delaying complex information was often an hindrance to the network.

The second example is taken from simulations of the balance beam task (a classic task in the developmental literature) first reported by McClelland & Jenkins (1991). These simulations occupy an important place in the book, because they show that a neural net using a continuous, non-linear learning function can display behaviour that looks stage-like and discontinuous. Unfortunately for this claim, however, Raijmakers, van Koten & Molenaar (1996) have recently shown that the stage-wise behaviour of the network is an artefact of the measures used by McClelland & Jenkins. When a proper set of measures is employed, with strict criteria for assessing discontinuity (criteria based on the mathematical catastrophe theory) it was shown that the neural net displayed no sign of stage-like behaviour.

In general, the simulations in Elman *et al*. come out best when they explore the dynamics of neural networks for themselves, without reference to empirical data. When they try to simulate data, they are plagued by the many problems noted by Rispoli. In this respect, it is unfortunate that Elman *et al*.

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

do not discuss alternative computational approaches, in particular adaptive production systems (e.g. Langley, 1987), which are, like neural nets, self-organizing systems displaying non-linear behaviour. In doing so, they could have contrasted both the strengths and weaknesses of neural nets, and perhaps found leads for further research. An obvious domain where this ecumenical approach would have helped is high-level cognition, such as reasoning, planning, or scientific thinking, about which connectionism is mostly silent, as noted by Elman *et al*. themselves (p. 394).

It is somewhat disappointing that Rispoli does not discuss the emphasis given by the book to recent advances in developmental neurobiology, because these constitute a cornerstone of Elman *et al*.'s argument. I believe that Elman *et al*. are right to challenge what they call representational innateness, which plays a prominent role in developmental psychology in general (e.g. the work of Carey and Spelke), and in developmental psycholinguistics in particular (e.g. the work of Chomsky and Pinker). Their detailed criticism of the data supposed to support innate representations of (cf. Chapter 7, where the matter is discussed with respect to language) and the evidence they adduce in favour of epigenesis (plasticity and equipotentiality of the brain, as well as the fact that the size of the human genome is not large enough to encode representational information) is compelling, although these questions are of course currently highly controversial in biology and in neuropsychology. Even so, their criticisms pose a welcome challenge to theoreticians who use 'explanations' based on innate factors, but do not provide detailed mechanisms.

Elman *et al*.'s argument becomes much weaker when they claim that connectionism offers a satisfactory explanation for epigenesis, however. Most of the models discussed in the book abstract too much from biological complexity to make this claim plausible. This is the case even in Chapter 6, which attempts to focus specifically on biologically plausible models. In particular, the simulations on the importance of starting small (Elman, 1993) are discussed at length in this chapter, as they are used to illustrate the importance of chronotopic constraints in the development of natural and artificial networks. As noted above, however, Rohde & Plaut (1997) were unable to replicate Elman's findings, showing that starting with full complexity leads to better results.

Elman *et al*. offer a nice introduction to non-linear dynamics, to neural nets, and to recent developments in developmental neurobiology. The attempt to put these three fields together into a single research framework does not come out convincingly, however. I think that this failure is due both to the complexity of these fields taken individually, and to the fact that our knowledge of the biological basis of development does not (yet) offer the constraints hoped for. Given that the simulations reported in the book are sometimes not replicable and often do not account for the data satisfactorily

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

– which both are of course obvious requirements for a model or theory that calls itself successful – I have to agree with Rispoli that Elman *et al*.'s book must be taken more as a position statement than as a scientific breakthrough.

REFERENCES

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition* **48**, 71–99.

Langley, P. (1987). A general theory of discrimination learning. In D. Klahr, P. Langley & R. Neches (eds), *Production system models of learning and development*. Cambridge, MA: MIT Press.

McClelland, J. L. & Jenkins, E. (1991). Nature, nurture and connections: implications of connectionist models for cognitive development. In K. VanLehn (ed.), *Architectures for intelligence*. Hillsdale, NJ: Erlbaum.

Raijmakers, M., van Koten, S. & Molenaar, P. (1996). On the validity of simulating stagewise development by means of PDP networks: application of catastrophe analysis and an experimental test of rule-like network performance. *Cognitive Science* **20**, 101–36.

Rohde, D. L. T. & Plaut, D. C. (1997). Simple recurrent networks and natural language: how important is starting small? *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (656–61). University of Stanford. Mahwah, NJ: Erlbaum.

Language acquisition also needs non-connectionist models

ULRIKE HAHN

University of Warwick

Rethinking innateness is a timely volume which forcefully demonstrates the importance of modelling in understanding development, 'innateness', and the nature of change. It provides an inspiring vision of what developmental psychology could one day be like, linking behaviour and biology via connectionist models. However, Rispoli's worry about the book's potential for detrimental polarization does not seem unfounded. One aspect of the book that deserves comment in this respect is the focus on connectionism to the exclusion of other types of model. It is unclear from *Rethinking innateness* itself whether this exclusion of other approaches merely stems from the legitimate desire to write a focused book, or whether it possibly reflects an actively held view that connectionism is the one true approach to modelling development. Regardless of the authors' intentions, the role of connectionism in relation to other computational approaches is an issue which is particularly pertinent to the study of language acquisition.

To clarify straightaway, I not only concur with Elman *et al*. on the central role of modelling, but also strongly believe that the task of the cognitive scientist is not complete until one has an account of how a particular process

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

is realised in a neural architecture. These two commitments give connectionist models a central role. But they do not make connectionism the exclusive modelling tool for the study of language acquisition, nor necessarily the best path to currently pursue. The reasons for this are twofold.

The first is a matter of research strategy. It is possible that more rapid progress and greater success might be made if one starts with high-level models which give less immediate regard to matters of implementation.

Promising directions might simply be easier to see with a top-down research strategy which starts from high-level task decompositions rather than with a strategy whose primary commitment is to start directly with a particular set of nuts and bolts. This holds especially because connectionism provides such a general framework and, in that sense, minimal constraint on the search space of possible models. Taking familiar architectures as a starting point in order to reduce the search problem could, for some aspects of language, turn out to be more of a hindrance than a help. For many areas of language, the currently popular backpropagation networks, which clearly constitute only a small segment of what would fit under the term 'connectionism', do not look like particularly good places to start. The central difficulty with respect to language is a whole host of representational issues. These include not only well-known problems with structured representations – which, to many, seem inevitable for syntax – but also a list of linguistically relevant factors, for instance (realistic) sentence semantics and discourse context, for which no obvious representation scheme is in sight. Another frequent difficulty is scaling problems, which can limit modelling to restricted data sets. Given that approaches which eschew 'innate representations' will tend to emphasize the role of domain inherent structure, an inability to make use of sufficiently large samples of real language (thus possibly distorting the structure in question) is a considerable worry.

These problems affect the study of different aspects of language to different extents. For instance, connectionist simulations with large-scale datasets have been possible for morphology (Hahn & Nakisa, 1996; Hahn, Nakisa & Plunkett, 1997). In general, it is no accident that connectionist models of inflectional morphology have infiltrated theorizing across the board in a way that models of syntax have not, and Rispoli's commentary provides a list of questions which, to date, lack significant connectionist contributions.

Accordingly, there may be areas where dissatisfaction with 'traditional' theorizing might better take the route of alternative models specified at a level well above even vaguely neural implementation. For instance, the probabilistic models and tools currently emerging within corpus linguistics suggest exploration by language acquisition theorists. Work here is typically symbolic but nevertheless deviates strongly from other traditionally held assumptions about the nature of linguistic knowledge, processing and acquisition.

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

This is no way to deny that neurally plausible implementation constrains the algorithmic solutions possible, with the consequence that moving from high-level description to neural implementation can never be a one-way street. The available building blocks clearly constrain the what and how of a design. But modifying a plan according to the available building blocks might sometimes be simpler than taking the blocks and trying to find a plan.

The second reason for welcoming a diversity of modelling approaches comes from considering the enterprise of cognitive science, and its approach to cognitive development, as such. Understanding neural processing AS computational (in contrast to a merely descriptive account of what brains happen to do) necessarily implies a general theoretical and empirical framework, i.e. what we know about computation. This invites model comparisons, and not just between connectionist models. Non-connectionist forms of computation provide an essential wider framework analogous to comparative physiology, comparative linguistics or anthropology. The utility of such frameworks stems from the fact that explanation should be provided in terms of general properties. This necessarily goes beyond merely providing a specific implementation. Generality uncontroversially demands multiple simulations, varying factors such as different initial weight settings or number of hidden units in order to demonstrate robustness of results. This can and should be expanded to variations of input and output representation and training procedures. Wider variations in network architecture are of interest too. All of this serves to establish the most general properties of networks which give rise to the desired behaviour given the problem at hand. Similarly, variations of the problem serve to elucidate its general aspects (see, for instance, Plunkett & Marchman, 1993). There is no reason, however, to stop here. Networks ought also to be compared with non-connectionist models. Statements about classes of learners and classes of problems are desirable. Only this kind of breadth allows one to identify both the general and the specific properties of (biological) neural architecture as a computational device.

Thus, even the most committed connectionist must acknowledge the potential contribution of non-connectionist approaches to language acquisition (as documented, for instance, in Brent, 1996). This extends to the potential contribution of the constantly evolving fields of machine learning, which is characterized by empirical comparisons of different learning algorithms (connectionist networks are frequently among these) and computational learning theory, which has not stopped with Gold.

In summary, connectionist modelling might well, for some areas of language and language acquisition, not be the fastest way forward and a diversity of modelling approaches will further our understanding of neural processes in the wider context of computation. Both seem inoffensive claims; their consequence is that, even for researchers deeply committed to bio-

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

logically plausible and realizable explanations, connectionism is not the only tool for the study of language development. Even those who share Elman *et al.*'s vision of what developmental psychology should become must look beyond connectionism.

REFERENCES

Brent, M. R. (1996). Advances in the computational study of language acquisition. *Cognition* **61**, 1–38.

Hahn, U., Nakisa, R. C. & Plunkett, K. (1997). The dual-route model of the English past tense: another case where defaults don't help. Proceedings of the 1997 GALA Conference on Language Acquisition. University of Edinburgh.

Nakisa, R. C. & Hahn, U. (1996). Where defaults don't help: the case of the German plural system. Proceedings of the 18th Annual Meeting of the Cognitive Science Society. Mahwah, NJ: Erlbaum.

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

Rethinking learning: comments on *Rethinking innateness**

VIRGINIA VALIAN

Hunter College and CUNY Graduate Center

In his review of *Rethinking innateness* (Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996; henceforth, RI), Rispoli (this volume) comments favourably on the dynamical change models presented in RI's Chapter 4. I think a more critical stance is warranted. In particular, I will argue that dynamical change models cannot in principle make reference to mental representation, that the models are stipulative, and that they fail as descriptions of behaviour. (For more extensive discussion, see Valian, in press.) The relation between dynamical change models and connectionist networks is not spelled out in RI, so it is not clear which of the criticisms that I direct at dynamical change models also hold for connectionist models.

Absence of concepts. The emphasis of RI's dynamical change models is continuity, a different kind of continuity from that proposed by nativists, which is continuity of concepts and theoretical vocabulary; nativism says nothing directly about mechanism. In RI, continuity refers to continuity of

[*] This research was supported in part by a grant from the National Institutes of Mental Health (MH 55353) and in part by a grant from The City University of New York PSC-CUNY Research Award Program. I thank Gary Marcus, Jerrold J. Katz, Fred Katz, Thomas Bever, Martin Chodorow, and Mary C. Potter for discussion and comments. I am also grateful to the authors of RI for critical comments of an earlier draft.

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

mechanism. New mechanisms do not come in; old mechanisms do not die out. Concepts, if they exist at all, can emerge *de novo*.

But are there any concepts? The general form of the dynamical change models in RI is (E). (More complicated models are discussed, but they do not change the basic points discussed below.)

$$(E) \quad dy/dt = by + c,$$

dy/dt refers to the change in a variable y over time t. y is always a performance measure, such as size of productive vocabulary in words or percent correct on a grammaticality task. Thus, while y is a behavioural consequence of knowledge, it is not itself knowledge. b and c are constants that represent mechanisms, such as learning ability and learning efficiency. The equations model behaviour by referring to mental mechanisms, but not to mental content or knowledge

Since we can interpret the terms of (E) very broadly, it might seem that y could stand for knowledge. If so, (E) could model the acquisition of knowledge. But for y to stand for knowledge, it would have to be reducible to values on a single dimension, like number of words in one's productive vocabulary. Lexical knowledge, however, is a congeries of different types of knowledge – phonological, syntactic, semantic, pragmatic. Even within word meaning there are different aspects of knowledge, ranging from the semantic domain the word is part of to specific aspects that differentiate one word from another. Operational definitions of lexical knowledge might be represented on a single dimension, but lexical knowledge itself cannot be.

Thus, equations like (E) are restricted to modelling behaviour and are at best only very indirectly related to the underlying abstract concepts to which the behaviour is related. Some dynamical change theorists have accepted the implications of that, saying, for instance, that 'knowledge ... is not a thing, but a continuous process; not a structure, but an action, embedded in, and derived from, a history of actions' (Thelen & Smith, 1994, p. 248). Dynamical change models of development, by their nature, are uninformative about mental structure. This is a serious problem, since things like the vocabulary spurt are of interest only because of what they might tell us about the organization of mental concepts and how learning takes place.

Since RI does not develop the connection between dynamical change models and connectionist nets, its position on knowledge is hard to ascertain. When it presents a connectionist model late in chapter 4, for example, it likens stages in Karmiloff-Smith's (1992) model of cognitive development – in which knowledge undergoes representational redescription – to stages in a connectionist model that detects the difference between odd and even numbers (p. 231). Here, then, they seem to intend to model knowledge and changes in knowledge. But I don't think the analogy holds up. In Karmiloff-Smith's model the child's underlying internal representation undergoes

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

qualitative change. In the recurrent network the behaviour simply looks as if there is a changing internal representation underneath it. Indeed, the whole point of the exercise is to demonstrate that nothing intrinsic changes except the interaction among mechanisms. As RI puts it, 'The question is, what sort of mechanisms might be responsible for what seem to be qualitatively different sorts of knowledge, and how can we move from one phase to the next?' (p. 231).

The example reflects a general inconsistency in the book's stance toward knowledge. On the one hand, it claims connectionist nets model the development of knowledge and show how new concepts, such as syntactic categories, can emerge. On the other hand, it claims that connectionist nets show that behaviour can look as if there are emergent concepts when in fact there is only learning of distributional regularities among contexts that correspond to entities that we label as syntactic categories. (See e.g. the discussion in Chapter 2.)

Stipulation. To return to dynamical change models like (E): in vocabulary development, the two mechanisms *b* and *c* can be thought of as ability to learn new words and learning efficiency. (Actually, the constants are interpreted differently in different places; I've picked what seems the most reasonable interpretation.) But calling *c* learning ability is unmotivated. RI has no theory of vocabulary development to which the constants in (E) are related. RI is not formalizing an existing theory with (E). There is no theory. The constants in (E) are merely the numbers that are needed to generate a curve. The numbers are named *ad hoc*. *c* is called 'ability to learn', but could be called anything.

Even if we assume that 'ability to learn' is the right name, we face the problem of stipulation. (E) simply stipulates that there is a single ability that remains constant, shedding no light on what that ability is. (E) treats as primitives the processes to be explained and described thus begging the developmental question.

The virtue of existing hypotheses for the vocabulary spurt (like the 'naming insight') is that they are explicit and specific enough to be tested and shown to be incorrect. The only explicit and specific components of RI's change models are the dependent variables that need to be accounted for in the first place. We could not test whether learning ability is a constant because we have no idea what learning ability is or how to measure it. That problem is hardly RI's alone. But RI treats learning ability as if we could take for granted that it is a univocal constant process.

A hypothetical example shows the difficulty. Someone who goes into a coma and then comes out of it really does experience a different rate of learning: there is a real step function down, followed by flat learning, followed by a step function up. RI could model the data as a dynamical change function by smoothing the steps out into a *U*. In this example we

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

know that the true story includes a second process. In vocabulary development, second language learning, and birdsong learning, we don't know what the true story is. Exactly what is at issue is the identity of the processes and their time courses. RI says it's possible that the processes are continuous. True enough. It's also possible that they aren't. It's possible that there are five different processes that enter and exit at different times. A possibility is not enough to lead us to prefer one model over another.

Curve-fitting. In principle, the choice of appropriate constants and values of constants will allow one to model the time course of almost any phenomenon. But the family of curves RI uses to fit data that resemble a step function smoothes out the function. There is no longer any 'step', as there would be in the coma example and as graphs like Figs. 4.7 and 4.9 show. On what grounds, then, would one prefer a curve that is so much more removed from the data? RI's answer, for both second language learning and birdsong learning, is that the dynamical change curve is a better empirical fit: it accounts for more of the variance in the phenomenon.

Does it? Take the data for second language learning (Johnson & Newport, 1989). Up until about age 16, second language learners' scores on a grammaticality judgement task decline with age of learning onset in an orderly, linear-looking way. In contrast, between about 16 and 40, no correlation is found; score variance also increases enormously. By eye, there is a downward sloping line and then an almost horizontal line. There appears to be something like a sensitive period for language learning, ending around age 16 (Johnson & Newport, 1989).

How much of the variance in language learning between birth and age 40 is accounted for by those two lines? RI says 39%, a figure obtained by averaging the variance accounted for in each of the two lines. The dynamical change curve accounts for 63% of the variance. But visual inspection of the different curves shows that RI's curve doesn't account for the variance in the first part of the data nearly as well as Johnson & Newport's (1989) line. And it does no better at accounting for the variance in the second part. How, then, could it account for more of the variance? It couldn't. The calculation of total variance allows the RI curve to profit from the correlation in the initial segment; the calculation of average variance prevents Johnson & Newport's two curves from profiting from its high initial correlation. What should be done is to total the squared deviation of each point from its respective line. In that case, the Johnson & Newport pair of lines would account for more of the variance. RI also reproduces data on birdsong learning (Marler & Peters, 1988). The empirical data show percentage of correct song learning flat at about 60% for the first 75 days or so of bird life. At that point there is a dramatic drop in correct learning, with the percentage varying between about 20% and 5% but having a downward slope. As with the second language learning data, there is also an increase in variance.

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

The curve generated by the dynamical change model wipes out the step function, thus missing the empirical phenomenon altogether. The birdsong data are the closest thing in nature that we have to the hypothetical example of a person entering a coma. Yes, we can model it as dynamical change and claim that the step function is a mirage. But why would we want to do that? Don't we instead want to find out whether what looks like a change is indeed a change that results in inability to learn?

I am not claiming that there is a sensitive period for language learning or birdsong learning, but that the discontinuities in the data cannot be eliminated by curve-fitting. The particular equations RI uses sacrifice the phenomenon that needs explanation. By producing a curve that wipes the phenomenon out we go backwards rather than forwards, because we make it seem as if there is nothing to explain.

Simplicity. RI suggests that we should prefer dynamical change models on the grounds of simplicity: they don't add or change mechanisms, while hypotheses like the critical period hypothesis for language or birdsong learning do. But two models can only be compared in simplicity if they account for exactly the same set of phenomena. That is not the case here. I have already noted that the dynamical change models do not account for the step function but eliminate it.

Even if that were not so, there is the problem of accounting for a cluster of phenomena. As RI notes, in the second language learning data there is an increase in variance at the same point that grammaticality judgements stop correlating with age of exposure. The dynamical change model does not account for that cluster. So it cannot be meaningfully compared with a sensitive period hypothesis; it does not account for the same data. The data reported for birdsong learning also show increased variance at the end of a hypothesized sensitive period, suggesting some generality to the phenomenon.

One possibility is that learning is tightly controlled by a dedicated mechanism until the end of the critical period. In language learning that mechanism appears to be less and less robust with age, while in birdsong learning the mechanism is equally strong throughout the critical period. In both cases, variance among learners is limited because the language-learning mechanism is dominating behaviour. When the critical period ends, the dedicated mechanism plays either a very limited role or no role; other, more general, learning mechanisms come into play. Those other mechanisms are both less efficient and more variable.

Is this story on the right track? We don't know, but we can find out. In humans, for example, we would expect post- but not pre-sensitive period behaviour to correlate with measures of cognitive efficiency. There are, of course, other approaches to the same phenomena. In second language learning, the mechanism of language learning may stay the same but

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

accessibility of linguistic concepts may decline. That too is something we can test. By looking at the entirety of a phenomenon we can get hints about what models could explain it.

Because dynamical change equations model only one behaviour at a time, they are particularly unsuited for describing domains like language development where phenomena are linked (see Marcus, in press, for a similar point). Thus, such models oversimplify and mischaracterize the phenomena to be explained.

Conclusion. Dynamical change models of language acquisition tell us nothing about the contents of the mind, do not fit the data, and are not simpler than other models.

REFERENCES

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.

Johnson, J. S. & Newport, E. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* **21**, 60–99.

Karmiloff-Smith, A. (1992). *Beyond modularity : a developmental perspective on cognitive science*. Cambridge, MA: MIT Press/Bradford Books.

Marcus, G. (in press). Can connectionism save constructivism? *Cognition*.

Marler, P. & Peters, S. (1988). Sensitive periods for song acquisition from tape recordings and live tutors in the swamp sparrow *Melospiza georgiana*. *Ethology* **77**, 76–84.

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

Valian, V. (in press). *Input and innateness : controversies in language acquisition*. Cambridge, MA: MIT Press.

Understanding the modelling endeavour

KIM PLUNKETT
Oxford University

JEFFREY L. ELMAN AND ELIZABETH BATES
University of California, San Diego

We enjoyed reading Matthew Rispoli's review of *Rethinking innateness* (henceforth RI-Elman *et al*. 1996). First, we respond to what we see as the major issues that he raises. Finally, we respond to the commentaries contributed by other authors in this issue.

Rispoli has two major concerns. First, he fears that RI will do the field a great disservice, increasing polarization and sowing discontent. To some extent this is an 'eye of the beholder' criticism that is difficult to counter: if Rispoli feels more polarized upon reading this book, who are we to argue?

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

We can only underscore that our intent is quite the opposite. It is, we think, an objective fact that the fields that study language are polarized around the issues of innateness and domain-specificity. We acknowledge this situation, and in the five years that we worked together to write the book, we were constantly aware of the need to replace heat with light, to work toward peace by working toward clarity. In our view, the current debate can be resolved, and invective can be replaced by an honest resolution, but this will only occur if we are clear in our goals, methods and terminology, providing operational definitions that permit an objective empirical test. For example, the taxonomy of 'ways to be innate' that we propose in Chapter 1 and illustrate throughout the book is presented as an opening bid, a proposal that encourages counter-proposals that are equally concrete and testable. If you say that 'X is innate', what do you mean, how could it be implemented, and what would it mean (in empirical terms) if you were right and we were wrong?

At some point in our history, all of the authors of this book have published joint works with individuals on the other side of the debate (and/or we ourselves were on the other side of the debate). Progress has been (and will be made) when we are clear about our differences and the differential predictions that result when we work together to solve a problem. This is much the same as the conclusion reached by MacWhinney in his commentary, while Maratsos notes that polarization is the rule rather than the exception in our field and concludes that while this may, at times, be unpleasant, it can in the long run lead to a theoretical synthesis.

The second major criticism concerns our approach to modelling. Rispoli argues that we eliminate a great deal from our models and in so doing make the results inevitable in a trivial way. In our view, building a good model is like designing a good experiment (see Plunkett & Elman, 1997, Chapter 2). One eliminates (or counterbalances) those factors that are irrelevant to a hypothesis. This does not entail a commitment to the view that eliminated factors are uninteresting, simply that they are not crucial to understanding the specific behaviour in question.

For example, in the past tense simulations, 'All the work of sorting through the verb paradigm and arriving at one form that means ''past tense'' has been done for the simulation'. In fact, Rispoli seems to think that we eliminate so much that the model becomes sterile. Suppose we had followed Rispoli's advice and incorporated additional sources of information such as 'semantic representations, grammatical feature, and paradigm, all of which are developmentally *a priori* to the sorting out of allomorphic variation'. Would it have behaved any differently? Rispoli's intuition seems to be that further complicating the model would make the task harder. We have two responses to this criticism, based on experience with models of this kind.

First, the behaviour may be more complex, but the essential characteristics often remain the same when additional factors are included (see, for example,

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

the Plunkett & Juola (in press) model of both English past tense and plural morphology). Second, adding complications and further constraints often has the paradoxical result of making it easier to learn. For example, adding semantic information to the past tense problem (as in MacWhinney & Leinbach, 1991; Cottrell & Plunkett, 1994) enables the network to solve problems that it cannot solve on the basis of phonological information alone. Complexity can work for you in surprising ways, especially when it reflects the contingencies found in the environment (Plunkett & Marchman, 1991).

Linguistics and the cognitive sciences more generally are replete with theories and experiments that voluntarily restrict their scope in the name of clarity and precision. Everybody, including those working with naturalistic free speech data, has to make hard decisions about what dimensions to include/exclude in a given study (Ochs, 1979). Connectionist modelling is no exception. Of course, there is no final excuse for failing to 'build a house from the foundations up'. However, we make no apologies for failing to build the house all at once. Often, it's useful to understand the properties of the building components before final construction commences.

In a similar vein, Rispoli argues that the Elman (1990) simulation is guilty of oversimplification. He supposes that nouns and verbs in that simulation are non-overlapping input categories. It is therefore no surprise that the network's representations of these categories are non-overlapping. He argues that introducing homophonic nouns and verbs into the training set would compromise the network's ability to represent word class information. This conclusion is incorrect.

Of course, input representations do influence a network's behaviour. Apparently, Rispoli misunderstood the nature of the representations used in the Elman (1990) simulations. In those models, all words had input representations that were orthogonal to each other, i.e. not only did nouns and verbs not overlap but any individual noun was no more similar to another noun than it was to any verb. That's why the distributional characteristics are all the network has to work with. So if homophones (identical input vectors which could occur in verb or noun positions) are included in the network training set (Rispoli is right, they were not), the network will respond on the basis of the form's position in the sentence, not its content. This is reflected in the network's capacity to generalize to novel forms. When the model sees a novel noun (an orthogonal input vector in noun position), it is treated as a noun, i.e. its internal representation is grouped together with the internal representations of nouns in the training set. 'Smell' is not a problem for this network.

Rispoli is equally worried about the absence of pronouns in the Elman (1990) simulation. For some reason, he believes that the simulation achieves its goal by avoiding pronouns, just as it purportedly achieved its goal by avoiding homophones. Not so. Let's suppose that Elman were to include

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

pronouns in his training set. Furthermore, suppose that he removes any surface information about the relation between the nominative and accusative forms of the pronoun by using orthogonal vectors to represent, say, 'him' and 'he'. Although the network has only got distributional information to go on, it will discover the relation between 'he' and 'him' as reflected in the similarity of the internal representations of these nominative and accusative forms. In fact, the similarity of the internal representations of the two pronouns will reflect that one is used as the subject of the sentence and the other is used as the object of the sentence. That the network can solve this kind of problem is demonstrated by its ability to capture the distinction between subject and object roles for the more difficult case of individual nouns. For example, the noun 'John' has a slightly different representation when used as a subject than when used as an object (Elman 1995). This difference is systematic across all nouns that can appear as subject and object. Learning this difference will be even easier when the relation is between accusative and nominative forms of distinct pronouns.

Do you need overt relative pronouns to achieve long distance agreement (where subject and main verb are separated by a subordinate clause)? Apparently not, as Christiansen & Chater (in press) have demonstrated. In addition, they show that these networks can process cross-serial dependencies, structures known to be computationally challenging. Ellis' commentary also seems to concur with our rebuttal of the apparent shortcomings pointed out by Rispoli, as does Chater's – 'Naturally, the closer the approximation of the input to that received by the child, the stronger the conclusions that can be drawn. But drawing negative conclusions on the basis of over-simplified input is unjustified' (Chater, this volume).

Rispoli faults us for failing to cite connectionist models that offer better approximations of reality. In fact, his examples were taken from works that were published after RI. However, we can only be pleased that the connectionist framework continues to make further progress and hardly see this as a criticism of our approach. Other commentators point to alternative frameworks for investigating the issues raised in RI. For example, Chater highlights the utility of statistical approaches in general; Ellis finds it unfortunate that we ventured all our eggs in one basket (connectionist simulations primarily done by Elman and Plunkett – in fact, a swift perusal of the bibliography reveals that we refer to many other connectionist models); Feldman laments our failure to treat 'localist' or 'structured' connectionist models in any detail; both Gobet and Hahn suggest that a broader sweep of computational approaches to development is needed. Again, we do not suggest that PDP models are the only appropriate computational tool for understanding development. However, we do claim that the connectionist models highlighted in RI offer important new insights into the nature of mechanisms underlying the developmental process, by

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

providing simulations of developmental trajectories and offering testable experimental predictions. These are issues all too often side-stepped in other computational approaches to development.

One final point, just a quibble really. It is difficult to fathom what version of Move Alpha Rispoli accuses us of badly misrepresenting. We are certainly not alone in describing Move Alpha as 'move anything, anywhere' (see Crain, 1991, p. 598). The whole point of this general rule is to reduce the number of rules of universal grammar.

Let us now turn to some specific issues raised by other commentators in this issue. Valian is concerned about the absence of concepts in RI. Focusing on Chapter 4, she points out that dynamic equations do not refer to mental content or knowledge. In fact, Chapter 4 is our attempt to relate the dynamical systems approach to connectionist modelling. Mental concepts and conceptual development are to be found throughout the book in the descriptions of the connectionist models themselves. These are replete with internal representations (admittedly, non-symbolic concepts) and theoretical assumptions about the conceptual apparatus and environment that are necessary to get learning off the ground.

Valian portrays the authors of RI as claiming that concepts emerge *de novo*. Not in any connectionist network we know about. Indeed, the message of RI is to flesh out a formal approach to interactionism. Nothing emerges from a *tabula rasa*. We have scoured RI for Valian's claim that we suppose 'New mechanisms do not come in; old mechanisms do not die out' but without success. We do claim that new behaviours can emerge without the need for new mechanisms (for example, see the account of vocabulary development in Chapter 3) but we do not claim that new behaviours never involve new mechanisms or that all behaviours involve the same type of mechanism. For example, we argue that different types of problem domains can best be solved by different types of network architecture that get recruited as the organism develops (see the section on Ways to be Innate in Chapter 1). We simply do not recognize the static approach that Valian attributes to RI.

In addition to her general concern about concepts and representations in neural networks. Valian accuses us of misleading the reader in our reanalysis (Chapter 4) of the Johnson & Newport (J & N) data on age of acquisition and second language learning. Although her account sounds plausible on the surface, it reflects a serious misunderstanding of basic linear statistics. In our reanalysis, we note that J & N broke their data on age of acquisition and linguistic performance down into two phases, before and after 16 years of age. In the first phase, the relation between age and performance was $-0{\cdot}87$, accounting for $75{\cdot}5\,\%$ of the variance; in the second phase, the relation was $-0{\cdot}16$, explaining only $2{\cdot}56\,\%$ of the variance. They conclude from this that two separate mechanisms govern learning before and after the critical period. We showed instead that a single curve (with no arbitrary break) can account

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

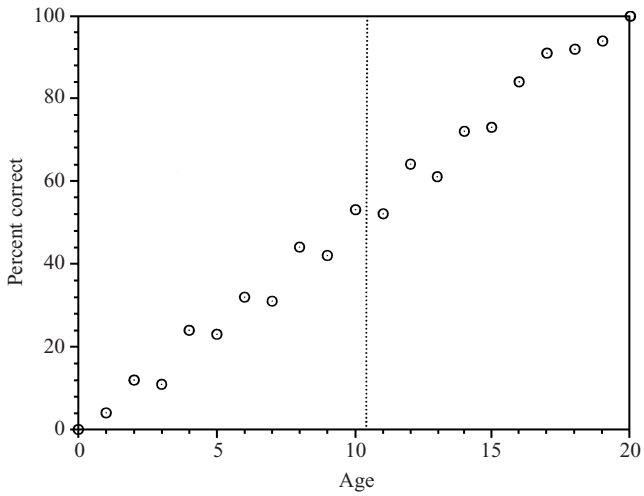


Fig. 3. An artificial data set illustrating possible correlations between age and performance.

for 63·1 % of the variance, whereas the two separate lines described by J & N account together for only 39·25 % of the variance (averaging the two data sets). We conclude that a single mechanism account explains more variance, and does so without postulating an arbitrary break. Valian suggests that our calculations are misleading: if one simply adds together the variance accounted for in J & N's Phase I and Phase II, then the J & N method accounts for 78·26 % of the variance, a better fit than we obtain with our single curve. But Valian's complaint is based on an error of logic. To illustrate this point, consider the artificial scatterplot in Fig. 3, representing a correlation of +0·995 between age and performance on some hypothetical measure.

This single linear relationship accounts for 99·1 % of the variance overall. Now, let's see what happens when we break the data down into two phases, before and after 10·5 years of age. In Phase I, the correlation is 0·979, accounting for 95·9 % of the variance; in Phase II, the correlation is 0·98, accounting for 96·6 % of the variance. If we average these two figures (as we did in Chapter 4), one obtains an estimate of 96·25 %, which is close to the total accounted for by a single line with no age breaks. However, if we follow Valian's recommendations and simply add the two variances together, we are forced to conclude that our two-phase model has accounted for 192·5 % of the variance in the relationship between age and performance! Valian suggests that there is no *a priori* reason to choose between her method and ours. However, as this example shows, our method is approximately correct while hers yields mathematically impossible results (i.e. no model can account for more than 100 % of the variance in the relationship between two measures).

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

Like Rispoli, Gobet argues that there are problems with various simulations reported in RI. For example, in discussing the 'starting small' simulation (Elman, 1993), Gobet claims that Rohde & Plaut (1997) failed to replicate the effect. In fact, the Rohde & Plaut simulations differed in several ways from those reported in Elman (1993). If anything, some of those differences (e.g. addition of semantics) make the connectionist language acquisition story even more plausible; other differences make learning easier, and so 'starting small' was not necessary (hence of no benefit). It is also worth pointing out that a number of other researchers have reported similar benefits to the 'starting small regime' when there are long-distance dependencies of the sort modelled by Elman. In fact, we know of no simulations described in RI that are not replicable. Indeed, a number of them are offered as demonstration exercises in the companion volume to RI (see Plunkett & Elman, 1997).

Gobet finds it unfortunate that we chose the McClelland & Jenkins (1991) balance beam model to demonstrate stage-like and discontinuous behaviour. Of course, there are many other models described in RI that exhibit stage-like and discontinuous behaviour. Nevertheless, the fact that mathematically identifiable 'catastrophic' discontinuities are not apparent in the balance beam model does not undermine the utility of its development message: Beware of interpreting discontinuities in behaviour (be they catastrophic or otherwise) as reflecting discontinuities in the underlying mechanism. For that way is fraught with danger. We stand firmly behind this message as one of the important contributions of connectionist modelling to our understanding of development.

REFERENCES

Christiansen, M. H. & Chater, N. (in press). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* **0**, 000–000.

Cottrell, G. W. & Plunkett, K. (1994). Acquiring the mapping from meanings to sounds. *Connection Science* **6** (4), 379–412.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences* **14**, 597–612.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science* **14**, 179–211.

—— (1993). Learning and development in neural networks: the importance of starting small. *Cognition* **48** (1), 71–99.

—— (1995). Language as a dynamical system. In R. F. Port and T. van Gelder (eds), *Mind as motion: explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

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

MacWhinney, B. & Leinbach, J. (1991). Implementations are not conceptualizations: revising the verb learning model. *Cognition* **40**, 121–57.

McClelland, J. L. & Jenkins, E. (1991). Nature, nurture, and connectionism: implications of connectionist models for cognitive development. In K. van Lehn (ed.) *Architectures for intelligence: the twenty-second Carnegie Mellon Symposium on Cognition*, 41–73. Hillsdale, NJ: Erlbaum.

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press

Ochs, E. (1979). Transcription as theory. In E. Ochs & B. B. Schieffelin (eds), *Developmental pragmatics*. New York: Academic Press.
Plunkett, K. & Elman, J. L. (1997). *Exercises in rethinking innateness: a handbook for connectionist simulations*. Cambridge, MA: MIT Press.
Plunkett, K. & Juola, P. (in press). A connectionist model for English past tense and plural morphology. *Cognitive Science* **00**, 000–000.
Plunkett, K. & Marchman, V. (1991). U-Shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. *Cognition* **38**, 43–102.
Rohde, D. L. T. & Plaut, D. C. (1997). Simple recurrent networks and natural language: how important is starting small? P. Langley (ed.), *Proceedings of the Nineteenth Annual conference of the Cognitive Science Society*. Stanford University: Mahwah, NJ: Erlbaum.

https://doi.org/10.1017/S0305000998243741 Published online by Cambridge University Press