Book Review

Jeffrey L. Elman, Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett, eds., *Rethinking Innateness: A Connectionist Perspective on Development*, Neural Network Modeling and Connectionism Series, Cambridge, MA: MIT Press, 1996, xix + 447 pp., $45.00 (cloth), ISBN 0-262-05052-8.

and

Kim Plunkett and Jeffrey L. Elman, *Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations*, Cambridge, MA: MIT Press, 1997, xv + 313 pp., $40.00 (paper), ISBN 0-262-66105-5.

Rethinking Innateness and *Exercises in Rethinking Innateness* are the first two of three planned volumes offering a connectionist perspective on development. The third volume will provide a collection of papers on connectionist models of development.

1

Brilliant in conception and naturally viewed as a sequel to the Parallel Distributed Processing volumes (Rumelhart et al. 1986), *Rethinking Innateness*'s multiple authors have ample space to develop a clear, well-articulated connectionist theory of development. Alas, the work is in many respects disappointing.

Chapter 1 ("New Perspectives on Development") suggests that recent advances in the neurosciences and connectionist modeling have something important to say to issues of innateness and development. These advances provide a way of fleshing out the idea that development involves an interaction between an organism's innate endowment and its environment, and that these interactions take place at a number of distinct levels. Foremost among these levels are representations (realized in connectionist networks as patterns of activation, connections, and weights between nodes), architectures (realized in connectionist networks in the structure and function of nodes, circuits, and collections of circuits), and timing (the order in which developmental processes take place). Elman et al. suggest that, to a first approximation, there are no innate cortical representations. By and large, what is innate are various properties of neural architecture and the timing of development. In addition to broaching theses concerning innateness, there is a first pass over the so-called "modularity of mind" (cf., e.g., Fodor 1983).

Chapter 2 ("Why Connectionism?") is a familiar sort of review of the basics of neural-network computation. Chapter 3 ("Ontogenetic Development: A Connectionist Synthesis") builds on the interactionist ideas of Chapter 1 by proposing that development involves emergent properties. The idea is explained through use of analogies. The spherical shape of a soap bubble is an emergent property of the local requirements concerning surface tension. The hexagonal shape of cells in a beehive is supposed to be an emergent property of the juxtapositioning of spherical cells. The spandrels of St. Mark's in Venice are supposed to be an emergent feature of the use of circular arches. The idea is also fleshed out in a number of models, including models of vocabulary development, acquisition of the ability to produce the past tense of English verbs, and the child's sensitivity to events in the physical world. Chapter 4 ("The Shape of Change") describes a range of possible patterns of development: linear change, non-linear monotonic change with linear dynamics, non-linear monotonic change with non-linear dynamics, and true discontinuous change. It also relates these types of change to dynamical systems and neural networks.

Chapter 5 ("Brain Development") discusses brain development at three levels of generality. First, there is a discussion of basic issues concerning the nature of development, including innateness, the putative modularity of mind, and a review of some findings in cellular neuroscience. Second, there is a review of the basics of vertebrate brain development and a survey of empirical results involving neural plasticity. The experiments on neural plasticity involve lesioning input pathways, regions of the brain, and transplanting plugs of neural tissue. The conclusion that Elman et al. draw from these studies is that brain tissue is not intrinsically dedicated to any one particular function, such as processing language, but can be recruited to serve a variety of distinct functions in other areas, such as vision and audition. The third section of Chapter 5 reviews some basic features of the human brain and what precautions one must bear in mind when trying to use animal data to make inferences about the human brain. This section also reviews a number of findings relating to brain plasticity in humans, cases of artificial and pathological lesioning.

Chapter 6 ("Interactions, All the Way Down") fills out a story begun in Chapter 1, namely, that development is a matter of interactions at all levels of analysis from the molecular to the whole organism. Unlike Chapter 1, however, the focus of Chapter 6 is on interactions among brain systems and between the whole organism and its environment. Three simulations describe how these interactions might take place within the connectionist framework. This chapter also makes a case for the view that an extended developmental process may be a biological adaptation, rather than a detrimental by-product of a conservative or "opportunistic" evolutionary process. A slow developmental process may enable a system like the brain to learn things it would not otherwise be able to learn. Development is, thus, supposed to pay for the investment of parental support in developing offspring and for the prolonged period of years in which the developing child is unable to fend for itself.

Chapter 7 ("Rethinking Innateness") turns to face the issues of innateness and the modularity of mind square on, addressing some of the supposed arguments for innate knowledge, including the species specificity of behavior, genetic disorders, localization of brain function, structural eccentricity, poverty of the stimulus, and the universality of a trait. There is also a list of future issues to be addressed by connectionist developmentalists.

A work like *Rethinking Innateness* has the potential to provide a clear, unified, well-argued extended statement of a connectionist perspective on development. It is the sort of work that might give greater direction to connectionist research on development. Yet it does not live up to all one might have hoped for. The work contains no substantially new technical innovations: no new weight change procedures, no new architectures, no new kinds of applications of existing architectures. Many discussions and analyses prove to be quite superficial: Obvious responses to certain arguments are unconsidered; alternative plausible interpretations of data are completely unaddressed. Further, the various chapters offer an inconsistent picture of how we are to understand innateness. The remainder of this review will elaborate on some of the more notable examples illustrating these difficulties.

Innate representations. One of the most contentious claims that Elman et al. make is that there are probably few, if any, innate cortical representations. The argument is that, since innate representations correspond to hard-wired patterns of weights or connectivity among nodes in a network, and since there appear to be no hard-wired patterns of weights or connectivity among neurons in the brain, there must be no innate knowledge (cf. pp. 25–26, 361). Yet, in over 400 pages of text, Elman et al. never consider the possibility that this argument only shows that representations should not be identified with patterns of weights and connections in a network.

Radical empiricism. Elman et al. claim that they are not radical empiricists, because, even though they reject representational nativism (the view that there is innate knowledge or innate representations), they allow that there may be innate determinants in the timing of developmental events and at the level of neural-network architecture (in the structure and function of nodes, circuits of nodes, and clusters of circuits of nodes). It is, however, a common assumption in cognitive science that a necessary (but not sufficient) condition on a process being cognitive is that it involve representations. Yet, if Elman et al. do not believe in innate representations, then it seems that they do not believe in any innate *cognitive* structure. But, if they do not believe in innate *cognitive* structure, how is it that they are not radical empiricists? Weren't the radical empiricists those who denied any innate *cognitive* structure? Doesn't being a nativist about the structure and function of nerve cells, and so forth, simply count as a kind of biological, but not psychological, nativism?

What is meant by 'innate'? Another, much more serious problem lies in figuring out exactly what we are supposed to believe about innateness after reading this book. It is clear that there is no consensus about this among the authors and that their diverse attitudes toward innateness surface in different chapters throughout the text. I count at least three different definitions of innateness, as well as two episodes challenging the very coherence or significance of the idea. In Chapter 1 (pp. 20–22), we find skepticism about the very concept of innateness:

> First, calling a behavior innate does very little to explain the mechanisms by which that behavior comes to be inevitable. So there is little explanatory power to the term. If all that is meant by saying a behavior is innate is that it is (under normal circumstances) inevitable, then we have gained little. (p. 21; cf. p. 357.)

There are, in addition, three definitions of innateness:

> An alternative definition is to reserve the term innate to refer to developmental outcomes that are more or less inevitable for a given species. That is, given the normal environment for a species, outcomes which are more or less invariant between individuals are innate. (p. 22; cf. p. 319.)

> Here the term innate refers to changes that arise as a result of interactions that occur within the organism itself during ontogeny. That is, interactions between the genes and their molecular and cellular environments without recourse to information from outside the organism. We adopt this working definition in this book. (p. 22; cf. p. 23.)

> To say that a behavior is innate is often taken to mean – in the extreme case – that there is a single genetic locus or set of genes which have the specific function of producing the behavior in question, and only that behavior. (p. 357.)

These definitions are clearly not equivalent. As Elman et al. point out (p. 23), a behavior can be universal in a species (given normal developmental conditions) without being endogenously determined. In the case of humans, normal developmental conditions lead essentially everyone to believe that people have feet. This would make the belief that people have feet innate under the first definition, but not the second. The second and third definitions also differ, since it is possible to have behavior that is endogenously determined through a complex set of genes and their interactions, rather than a single gene. The first and third definitions are obviously different as well.

Elman et al. do not indifferently shift between these definitions; they choose a definition so as to facilitate a particular result. In their criticism of representational nativists, in Chapter 7, Elman et al. suppose that the nativists hold the view that a behavior or characteristic is innate if there is a single genetic locus or set of genes that specifically codes for the behavior. This is the strongest of the three definitions of innateness given above and the one they attribute to the representational nativists without textual support. (The fact that there is no textual support for this particular attribution is striking, given the fact that Elman et al. specifically attempt to rebut the view that they are attacking a Straw Nativism; p. 367). Having made this attribution, Elman et al. go on to show how familiar psychological claims do not support the strong conception of innateness. Here they are in action dealing with the species specificity of language acquisition:

> Clearly no one denies that human beings are the only creatures that are able to learn and use grammar in its full-blown form. While studies of symbol use and language-like behavior in chimpanzees and other nonhuman primates are interesting, all of the animals studied to date stop far short of the language abilities displayed by a normal human child at 3 years of age. However, it does not automatically follow that grammar itself is under strict genetic control because, as we have shown repeatedly in this volume, behavioral outcomes can be the product of multiple levels of interaction in which there is no representational prespecification. (p. 372.)

So, the strong conception of nativism is unsupported. Here one would think that Elman et al. might at least consider an alternative understanding of what is meant by 'innateness' in this "species specificity" argument, but they do not.

In Chapter 6, Elman et al. discuss chicks imprinting on mother hens to make the point that "behaviors may be innate and yet nonetheless involve considerable interactions. Thus, interaction and innateness are not incompatible. Rather, we argue that interaction is an important way in which behaviors may be innate" (p. 322). In this example, there are interactions between brain systems in the chick, the so-called *Conspec* region and the IMHV region, but there are also interactions between the chicks and their environments. In order for the *Conspec* region to develop, a chick must be allowed to walk around in its environment in the period from 12 to 36 hours after hatching. This locomotion appears to stimulate the production of testosterone essential for the development of *Conspec*. In order for the chick then to imprint on a mother hen, it must attend to the mother hen. If this example of chick imprinting is meant to show that innateness and environmental interaction are compatible, 'innate' cannot mean "endogenously determined". (In other words, they cannot be using 'innate' in the way they say they will in Chapter 1, p. 22.) This explains why, in Chapter 6 (p. 319), Elman et al. take 'innate' to mean something like "highly probable (universal) given normal environmental conditions". Given this definition, however, it is trivial to draw the conclusion that interaction and innateness are compatible.

Although Elman et al. appear to secure the thesis that innateness and interaction are compatible by means of a very loose conception of innateness, one might think the compatibility thesis could be secured with a more plausible definition of innateness (one that Elman et al. do not introduce). Thus, a belief or behavior is innate if it is not learned on the basis of information in the environment. On such an analysis, the development of *Conspec* would be innate, yet would involve an interaction with the environment. The development of *Conspec* appears to involve an interaction with the environment, but this development does not appear to be an instance of learning, since *Conspec* does not gather information about the environment. It is merely the non-specific influence of testosterone that is needed. With this definition, one could obtain the slightly more interesting conclusion that innateness and interaction are compatible: Innateness and interaction with the environment are compatible as long as the organism's interaction with the environment is not an instance of gathering information from the environment.

Innateness and modularity. Another difficulty with the book is the apparent conflation of innateness with issues about domain specificity of cognitive processing, localization of function, and the modularity of mind. (Perhaps 'conflation' is too strong a word here; cf. p. 371.) The issue of innateness is relevant to the theory of development, where the theory of modularity has to do with the structure of the adult mind. Theses about innateness and the modularity of mind and domain specificity are, therefore, logically separable. Elman et al. write as if this were a point of contention, but in fact the logical separability of innateness from modularity is a crucial feature of Jerry Fodor's view of modularity. According to Fodor, input systems constitute a natural kind because they share many properties that are not logically implied by their definition (Fodor 1983, p. 46). Thus, input systems mediate between peripheral transducers and central cognitive systems involving beliefs and desires. Contrary to the suggestion by Elman et al., it is not part of the definition of input systems that they be informationally encapsulated, unconscious, fast, innate, have shallow outputs, etc. Rather, the logical independence of these concepts is part of Fodor's case for thinking that they are scientifically interesting. So, there is apparently no issue about the independence of domain specificity and innateness separating Fodor from Elman and his colleagues.

Rethinking Innateness is perhaps best viewed as a missed opportunity. The numerous ideas floating around in connectionist theorizing simply have not been neatly and coherently synthesized here. The real contribution of this book must, therefore, lie in the discussion of numerous smaller ideas and arguments. In this regard, there is much to be examined in *Rethinking Innateness*.

2

Exercises in Rethinking Innateness is the companion volume to *Rethinking Innateness*, containing the Tlearn neural-network simulator, version 1.0. This software runs a range of back-propagation-based neural-network simulations on the Macintosh and Windows 95 platforms. In this section, I will comment on both the text and the software, which I have used in an upper-division undergraduate course, Introduction to Neural Networks.

Chapter 1 ("Introduction and Overview") reviews the basic nuts and bolts of nodes, connections, weights, activation values, inputs, outputs, and so forth. Chapter 2 ("The Methodology of Simulations") briefly introduces the concept of a neural-network simulation and provides a rationale for using computer simulations. Chapter 3 ("Learning to Use the Simulator") introduces the reader to the basic features of the Tlearn simulator through two-layer networks for computing the boolean functions "and" and "or". It also examines the inability to handle "xor". Hinton diagrams, a system for visually presenting positive and negative weights of various magnitudes, are also introduced. Chapter 4 ("Learning Internal

Representations") examines how three-layer networks solve the "xor" problem. It also covers the Tlearn facilities for examining hidden node activations in response to input activations. Chapter 5 ("Autoassociation") examines autoassociation and its application to forming distributed representations from localist representations, finding redundancies in data sets, and performing pattern completion. It also introduces a technique for analyzing what hidden nodes represent. Chapter 6 ("Generalization") discusses the ability of networks to generalize, for the first time drawing a distinction between a training set and a testing set disjoint from the training set. The technique of cluster analysis is also discussed here.

Chapter 7 ("Translation Invariance") marks a subtle shift from more purely technical, computational properties of networks and their analysis to the deployment of connectionist networks as models of psychological processes. It explores an example of how one might use the receptive-field idea, found in Rumelhart, Hinton, and Williams 1986, to identify strings of three contiguous 1s in an input field. The model developed here has the ability to generalize from a training set to a test set in a way that a simple three-layer network is unlikely to do. Chapter 8 ("Simple Recurrent Networks") applies a simple recurrent network in a letter-prediction task. This model takes advantage of a Tlearn facility for coding strings of symbols, such as letters in the English alphabet, into strings of 1s and 0s that can be used by the network. Chapter 9 is a relatively perfunctory discussion of "Critical Points in Learning". Chapter 10 ("Modeling Stages in Cognitive Development") discusses James McClelland's (1989) balance-beam model and the concept of developmental phases. The explanation of the problem, its theoretical significance, and how the network is supposed to solve the problem is a bit on the short side. Chapter 11 ("Learning the English Past Tense") is an introduction to the debate over Classical and Connectionist models of learning the past-tense of English verbs. This chapter may be the most theoretically taxing for the novice, since it presupposes a fair amount of background regarding phonology, the types of past-tense verbs, and the U-shaped learning curve for past-tense verbs. Finally, Chapter 12 provides a rather dense introduction to "The Importance of Starting Small". The idea is that by beginning learning with a fragment of the ultimate training sets, a network can ultimately learn a larger training set.

The book also contains two appendices. The first explains the installation of Tlearn, a way to report bugs, and a procedure for obtaining Tlearn via anonymous ftp. The second, much larger, appendix is a good user's manual.

Each chapter in *Exercises* develops one or two principal ideas relating both to the theory of neural networks and the various facilities provided by Tlearn. Further, subsequent chapters build on preceding chapters in a logical way. Each chapter also contains exercises that develop or reinforce the user's understanding of the properties of neural networks. These exercises are then answered at the end of the chapter, often in detail. For example, part 2 of Exercise 3.3 explores how small the root-mean-square error must be in order to guarantee that a given network has solved a given problem. Part 1 of Exercise 3.7 draws attention to the significance of different initial weights. Exercise 4.7 explores the interaction between the learning rate and momentum in the backprop weight-change procedure. Exercise 6.7 explores the significance of varying the number of hidden nodes in a network. Exercise 8.1 asks why, in simple recurrent networks, context nodes are fully connected to the hidden nodes, when the hidden nodes are not fully connected to the context nodes. These exercises are quite good; however, this generally leaves the instructor to fashion homework exercises that will build up a student's comfort level in using the Tlearn software and neural networks. The instructor must develop the exercises that will enable the student to become proficient with the technical details of operating the Tlearn software. While this is not terribly problematic for an instructor, it might make the software more challenging for an individual working alone. So, additional exercises with this aim might improve the book.

Since this book appears to be aimed at the psychological-modeling community, another way in which it might be improved is by providing more numerous examples of the ways in which the models might be relevant to psychological phenomena. For example, in Chapter 5, we are told that autoassociation is useful for forming distributed representations from localist representations, finding redundancies in data sets, and performing pattern completion in the face of noisy input. How might this be related to psychological processes? Or, in Chapter 6, how does generalization in learning a binary-addition problem and the categorization of 4-bit vectors into two categories apply more widely? In Chapters 8 and 9, one might wonder how simple recurrent nets might be applied to psychological processes other than those examined in the text. The actual cases developed in the text are predicting the next letter in a sequence (Chapter 8) and reporting whether the network has seen an odd or even number of 1s (Chapter 9). A little more guidance in this area would probably be useful.

The later chapters of the text, where more advanced concepts are employed, are more demanding. The text's discussion and explanation of such things as cluster analysis, cross entropy, the use of eigenvectors, and lesioning vary from the merely perfunctory to the non-existent. While a course instructor might fill in here, there are no references to the primary literature that might aid an independent investigator. Some additional explanation would be of obvious value here.

Turning to the Tlearn software itself, the most important fact is that it is a back-propagation simulator. This limits the range of courses of study for which the text and software might be used. It is probably best for a cognitive-science course where the focus is on cognitive modeling, rather than a computer-science course where a familiarity with the diversity of neural networks and their computational properties may be of interest. Another feature of Tlearn also supports this use of the text and software: The system for configuring networks really only works well for completely connected two- and three-layer networks. Simple recurrent networks are not really a problem, but the grouping of connections into receptive fields proves to be rather tedious for all but the very simplest of cases. On the up side, Tlearn provides a nice range of facilities for analyzing networks and their behavior. One slight irritation lies in the facility for displaying the network architecture. All the hidden nodes in a network are displayed in a single row, making networks with more than three layers – such as a simple recurrent networks – less visually clear.

Although Tlearn uses a rather standard Windows interface, there remain a sufficient number of instances where file manipulation proves to be a challenge even for students who have some familiarity with Windows. In addition, Tlearn has a good supply of bugs, some of them quite serious. The most serious I've found involves the command to display the network architecture. Calling this function will at times lead Tlearn to crash. Worse yet, it sometimes consumes RAM in such a way as to prevent Tlearn from being restarted, essentially bringing down the whole system. Another less serious problem sometimes occurs when the program is asked to test a neural-network on novel data not used to train the network. This bug will cause Tlearn to fail to recognize crucial files in the particular neural-network project on which one is working. One must then, apparently, close all the files one is using and reopen them. There is also a facility that is supposed to encode representations such as strings of letters into strings of 1s and 0s of the sort a backprop network can take as input. The translation facility often introduces errors into the coding, which makes its use extremely problematic. Given these bugs with Tlearn 1.0, it is probably worth the user's time to obtain any updated versions that might be available via anonymous ftp.

While not without its shortcomings, *Exercises* is a good book for certain applications, such as an introductory neural-networks courses for psychological modeling. In addition, with the promised improvements to the software, one can expect it to become even more useful.

References

Fodor, Jerry A. (1983), *The Modularity of Mind*, Cambridge, MA: MIT Press.

McClelland, James L. (1989), 'Parallel Distributed Processing: Implications for Cognition and Development', in R. G. M. Morris, ed., *Parallel Distributed Processing: Implications for Psychology and Neurobiology*, Oxford: Clarendon Press, pp. 9–45.

Rumelhart, David; Hinton, Geoffrey; and Williams, Ronald (1986), 'Learning Internal Representations by Error Backpropagation', in David Rumelhart and James L. McClelland, eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1, Foundations*, Cambridge, MA: MIT Press, pp. 318–362.

Rumelhart, David; McClelland, James L.; and the PDP Research Group (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 2 vols., Cambridge, MA: MIT Press.

KENNETH AIZAWA

Department of Philosophy,
Centenary College of Louisiana,
Shreveport, LA 71134
E-mail: kaizawa@beta.centenary.edu