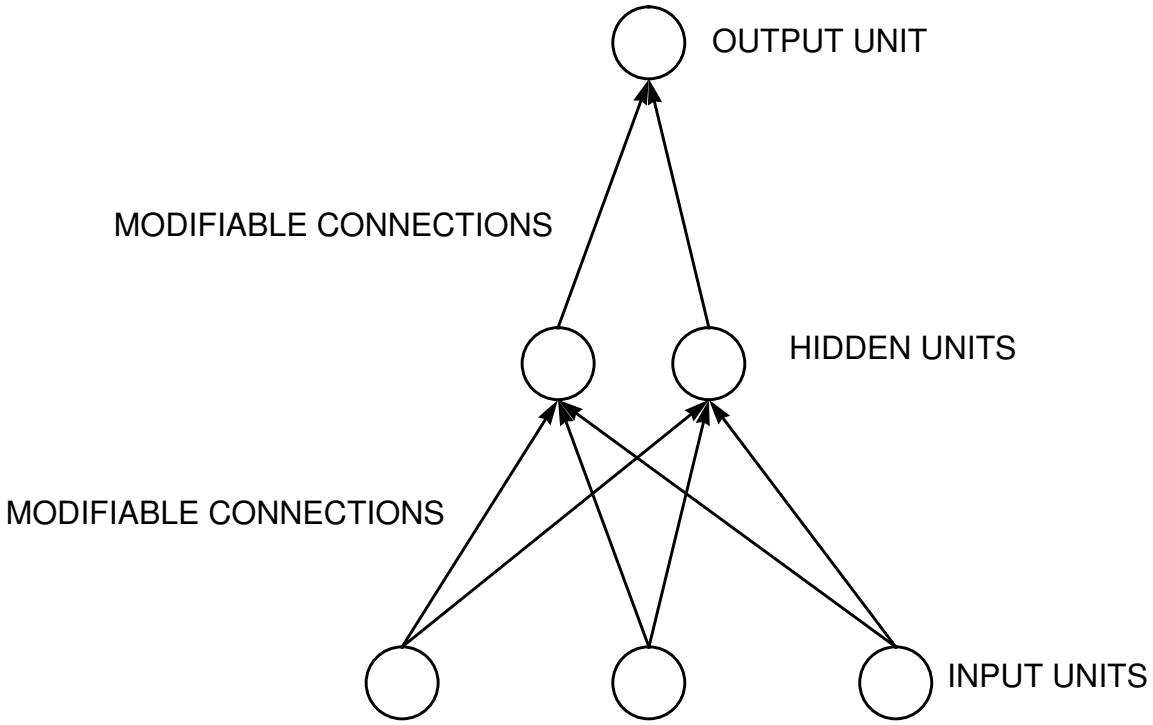

Figure 2

A multi-layer network. The "hidden units" allow such networks to form internal representations. Multi-layer networks of this sort have been shown to be able to approximate any function to any aribtrary degree of accuracy.

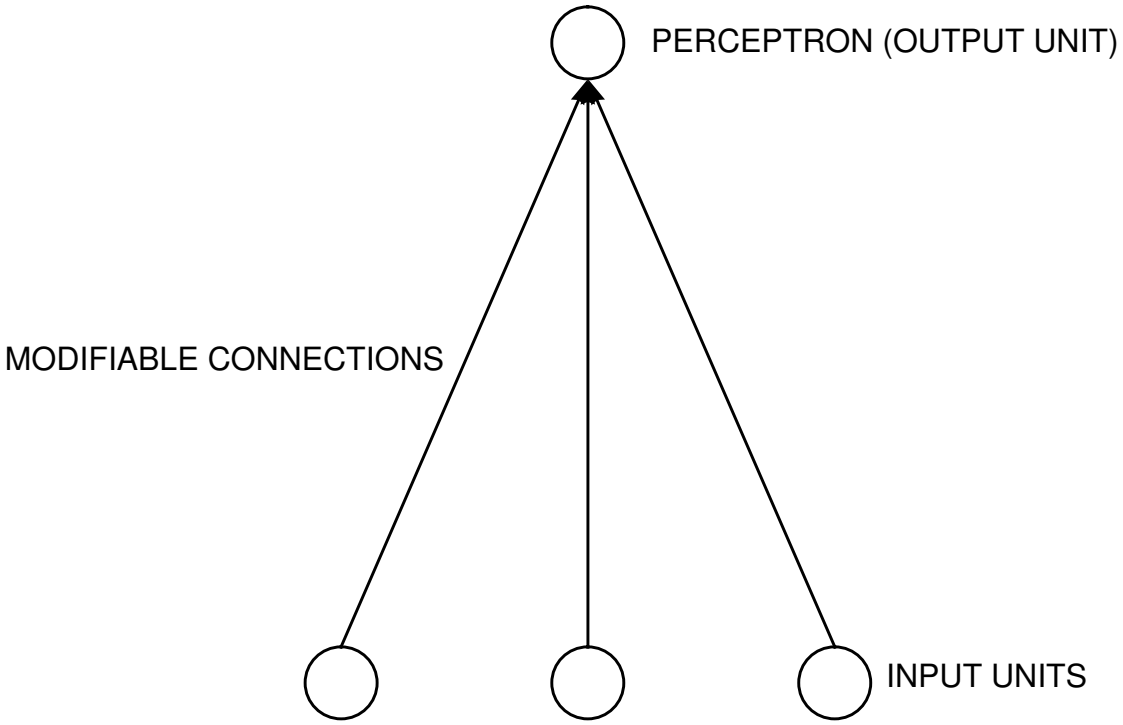

Figure 1

A simple perceptron network.

Early lexical development in children with focal brain injury. *Brain and Language*, **40**, 491-527.

von Neumann, J. (1951). The general and logical theory of automata. In L.A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley.

von Neumann, J. (1958). *The computer and the brain.* New Haven: Yale University Press.

Werner, H. (1948). *Comparative psychology of mental development*. New York: International Universities Press.

Willshaw, D.J., Buneman, O.P., & Longuet-Higgins, H.C. (1969). Nonholographic associative memory. *Nature*, **222**, 960-962.

Neurocomputing: Foundations of research. Cambridge, Mass.: MIT Press.

Mead, C. (1989). Analog VLSI and neural systems. Inaugural address presented to the Institute for Neural Computation, October, 1989. University of California, San Diego.

Minsky, M. (1956). Some universal elements for finite automata. In C.E. Shannon & J. McCarthy (Eds.), *Automata studies.* Princeton: Princeton University Press. Pp. 117-128.

Minsky, M. and Papert, S. (1969). *Perceptrons*. Cambridge, Mass.: MIT Press.

Papert, S. (1988). One AI or Many? *Daedalus: Artificial Intelligence.* Winter, 1988.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.

Piaget, J. (1970a). *Structuralism*. New York: Basic Books.

Piaget, J. (1970b). *Genetic epistemology*. New York: Columbia University Press.

Piaget, J. (1971). *Biology and knowledge: An essay on the relations between organic regulations and cognitive processes.* Chicago: University of Chicago Press.

Piatelli-Palmarini, M. (1989). Evolution, selection, and cognition: From "learning" to parameter setting in biology and the study of language. *Cognition,* **31**, 1-44.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. In S. Pinker & J. Mehler (Eds.), *Connections and Symbols.* Cambridge, MA: MIT Press/Bradford Books. Pp. 3-71.

Plunkett, K. and Marchman, V. (1991a) U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. *Cognition*, **38**, 43-102.

Plunkett, K. and Marchman, V. (1991b) From rote learing to system building. (Technical Report 9020). Center for Research in Language, University of California, San Diego.

Roeper, T. and Williams, E., Eds. (1987). *Parameter setting*. Dordrecht and Boston: Reidel.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context.* New York: Oxford University Press.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, **65**, 386-408.

Rosenblatt, F. (1962). *Principles of neurodynamics*. New York: Spartan.

Rumelhart, D., Hinton, G. and Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533 - 536.

Rumelhart, D., McClelland, J. and the PDP Research Group (1986). *Parallel distributed processing: explorations in the microstructure of cognition, Vol. 1*. Cambridge, Mass.: MIT/Bradford Books.

Schwartz, M.F., Saffran, E.M., & Dell, G.S. (1990). Comparing speech error patterns in normals and jargon aphasics: Methodological issues and theoretical implications. Presented to the Academy of Aphasia, Baltimore, MD.

Seidenberg, M., & McClelland, J.L. (1989). A distributed developmental model of visual word recognition and naming. Psychological Review, 96, 523-568.

Selfridge, O.G. (1958). Pandemonium: a paradigm for learning. In *Mechanisation of Thought Processes: Proceedings of a Symposium Held at the National Physical Laboratory, November 1958.* London: HMSO. Pp. 513-526.

Spelke, E. (1990). Principles of object perception. *Cognitive Science*, **14**, 29-56.

Spelke, E. (1991). Physical knowledge in infancy: Reflections on Piaget's theory. In S. Carey and R. Gelman (Eds.), *The epigenesis of mind: essays on biology and cognition*. Hillsdale, New Jersey: Erlbaum, 133 - 169.

Thal, D., Marchman, V., Stiles, J., Aram, D., Trauner, D., Nass, R., & Bates, E. (1991).

typological introduction. Volume I. Amsterdam: John Benjamins.

Grossberg, S. (1968). Some physiological and biochemical consequences of psychological postulates. *Proceedings of the National Academy of Science, USA*, **60**, 758 - 765.

Grossberg, S. (1972). Neural expectation: Cerebellar and retinal analogs of cells fired by leranable or unlearned pattern classes. *Kybernetik* 10, 49 - 57.

Grossberg, S. (1987). *The adaptive brain,* 2 vols. Amsterdam: Elsevier.

Hebb, D. (1949) *The organization of behavior.* New York: Wiley.

Hertz, J., Krogh, A. and Palmer, R. (1991). *Introduction to the theory of neural computation*. Redwood City, California: Addison Wesley.

Hinton, G.E., & Shallice, T. (1991). Lesioning a connectionist network: Investigations of acquired dyslexia. *Psychological Review*, **98**, 74-95.

Hinton, G.E., & Anderson, J.A. (1981). *Parallel models of associative memory.* Hillsdale, NJ: Erlbaum.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, **2**, 359-366.

Hyams, N. (1986). *Language acquisition and the theory of parameters*. Dordrecht & Boston: Reidel.

Jacobs, R., Jordan, M. and Barto, A. (1991). Task decomposition through competition in a modular connectionist architecture: the what and where visual tasks. *Cognitive Science*, **15**, 219 - 250.

Kellman, P. J.,Spelke, E. S., & Short, K. R. (1986). Infant perception of object unity from translatory motion in depth and vertical translation. *Child Development*, 57, 72-86.

Kirsh, D. (1992). PDP Learnability and innate knowledge of language. *Center for Research in Language Newsletter Vol. 6, no. 3.* University of California, San Diego.

Kohonen, T. (1977). *Associative memory: A system-theoretical approach.* Berlin: Springer.

Lachter, J., & Bever, T.G. (1988). The relation between linguistic structure and associative theories of language learning: A constructive critique of some connectionist learning models. In S. Pinker & J. Mehler (Eds.), *Connections and Symbols.* Cambridge, MA: MIT Press/Bradford Books. Pp. 3-71.

Lakoff, G. (1987). *Fire, women, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Langacker, R. (1987). *Foundations of cognitive grammar: Theoretical perspectives. Volume I.* Stanford: Stanford University Press.

Le Cun, Y. (1985). Une procédure d'apprentissage pour réseau à seuil assymétrique. In *Cognitiva 85: à la Frontière de l'Intelligence Artificielle des Sciences de la Connaissance des Neurosciences* (Paris 1985), 599 - 604.

Lightfoot, D. (1991). The child's trigger experience — Degree-0 learnability. *Behavioral Brain Sciences*, 14:2.

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

Marchman, V. (1992). Language learning in children and neural networks: Plasticity, capacity, and the critical period. (Technical Report 9201). Center for Research in Language, University of California, San Diego.

McClelland, J. and Rumelhart, D. (1986). *Parallel distributed processing: explorations in the microstructure of cognition, Vol. 2.* Cambridge, Mass.: MIT Press/Bradford Books.

McCulloch, W. and Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5, 115 - 133. Reprinted in J. Anderson and E. Rosenfeld (Eds.),

change.

FOOTNOTES

1. This particular school of Functionalism has little to do with, and is indeed diametrically opposed to, an approach within linguistics and psycholinguistics alternatively called Functional Grammar or Cognitive Linguistics. For discussions, see Bates and MacWhinney, 1989; Langacker, 1987; Lakoff, 1987; Givón, 1984.

2 A number of readable introductions to connectionism are now available. See Bechtel and Abrahamsen, 1991; Churchland and Sejnowski, in press; Dayhoff, 1990. An excellent but more technical introduction can be found in Hertz, Krogh, & Palmer, 1991.

REFERENCES

Anderson, J.A. (1972). A simple neural network generating an interactive memory. *Mathematical Bio-Sciences, **8***, 137-160.

Anderson, J.A., & Rosenfeld, E. (1989). *Neurocomputing: Foundations of Research.* MIT Press/Bradford Books.

Baillargeon, R., & de Vos, J. (1991). Object permanence in young infants: Further evidence. *Child Development*, **62**, 1227-1246.

Bates, E., Thal, D. and Marchman, V. (1991). Symbols and syntax: A Darwinian approach to language development. In N. Krasnegor, D. Rumbaugh, E. Schiefelbusch and M. Studdert-Kennedy (Eds.) *The biological and behavioral determinants of language development*. Hillsdale, NJ: Erlbaum.

Bruner, J., & Sherwood, V. (1976). Peekaboo and the learning of rule structures. In J. S. Bruner, A. Jolly & K. Sylva (Eds.), *Play: Its role in development and evolution*. New York: Basic Books, Inc.

Bechtel, W. and Abrahamsen, A. (1991). *Connectionism and the mind.* Oxford: Basic Blackwood.

Churchland, P. and Sejnowsky, T. (in press). *The net effect.* Cambridge, MA: MIT Press/Bradford Books.

Crick, F. (1989). The recent excitement about neural networks. *Nature*, **337**, 129 - 132.

Dayhoff, J. (1990.) *Neural network architectures.* New York: Van Nostrand Reinhold.

Eccles, J.L. (1953). *The neurophysiological basis of mind.* Oxford: Clarendon.

Elman, J. (1989). Structured representations and connectionist models. In *The Eleventh Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

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

Elman, J. (1991) Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, **7**, 195-225.

Feldman, J. A., & Ballard, D.H. (1980). Computing with connections. TR 72. University of Rochester: Computer Science Department.

Ferguson, C., & Snow, C. (1978). *Talking to children.* Cambridge: Cambridge University Press.

Fodor, J.A. (1981) *Representations*. Brighton (Sussex): Harvester Press.

Fodor, J.A., & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: A critical analysis. In S. Pinker & J. Mehler (Eds.), *Connections and Symbols.* Cambridge, MA: MIT Press/Bradford Books. Pp. 3-71.

Givón , T. (1984). *Syntax: A functional-*

researchers in the connectionist movement who are trying to bring these systems closer to neural reality. For example, efforts are underway to study the computational properties of different neuronal types. Some researchers are exploring analogues to synaptogenesis and synaptic pruning in neural nets. Others are looking into the computational analogues of neural transmitters within a fixed network structure. The current hope is that work at all these different levels will prove to be compatible, and that a unified theory of the mind and brain will someday emerge. Of course we are a long way off, but the commitment by most of the researchers that we know in this field is a very serious one. It has launched a new spirit of interdisciplinary research in cognitive neuroscience, one with important implications for developmental psychology.

Worry #5. "Connectionism is antinativist, and efforts are underway to reinstate a *tabula rasa* approach to mind and development" (e.g., Kirsh, 1992). It is true that many current simulations assume something like a *tabula rasa* in the first stages of learning (e.g., a random "seeding" of weights among fully-connected units before learning begins). This has proven to be a useful simplifying assumption, in order to learn something about the amount and type of structure that has to be assumed for a given type of learning to go through. But there is no logical incompatibility between connectionism and nativism. Indeed, just as many historians have argued that Franklin Delano Roosevelt saved capitalism, connectionism may prove to be the salvation of nativist approaches to mind. The problem with current nativist theories is that they offer no serious account of what it might mean in biological terms for a given structure or idea to be innate. In neural networks, it is possible to explore various avenues for building in innate structure, including minor biases that have major structural consequences across a range of environmental conditions (Jacobs, Jordan, & Barto, 1991). In fact, within connectionist models there are coherent ways to talk about 90% or 10% of any innate idea! This is an approach that has not been explored in any detail to date, but the possibilities are intriguing, and might (ironically enough) end up being connectionism's greatest contribution to developmental cognitive neuroscience.

To conclude, we are willing to speculate that we will soon see a revival of Piagetian theory within a connectionist framework — not a mindless reinterpretation of the old theory in modern jargon, but a return to Piaget's program of genetic epistemology, instantiating his principles of equilibration and adaptation in concrete systems that really work — and really change. As we said before, Piaget spent the later decades of his life seeking a way of formalizing the theory, to answer critics (including Piaget himself) who charged that his principles of change were much too vague. We think that Piaget would have loved these new possibilities if he had lived to see them. We now have an opportunity to pick up the threads of his old program and move it forward into an exciting new decade, incorporating all the new insights and new empirical information that has been gained in the interim, without abandoning the fundamental commitment of developmental psychology to the study of

ready know the limits of this kind of associationism has been misinformed.

Worry #2 . "There are no interesting internal representations in connectionist nets" (e.g., Pinker & Prince, 1988). There are indeed complex and rich representations in connectionist networks, and transformations that do the same work as rules in classical systems. However, these rules and representations take a radically different form from the familiar symbols and algorithms of serial digital computers and/or generative linguistics. The representations and rules embodied in connectionist nets are implicit and highly distributed. Part of the challenge of modern research on neural networks is to understand exactly what a net has learned after is has reached some criterion of performance. So far, the answer appears to be that they do not look like anything we have ever seen before (for examples, see Elman, 1989, 1990, 1991).

Worry #3 . "Connectionist nets only yield interesting performance on cognitive problems when the experiment 'sneaks in' the solution by (a) fixing the internal weights until they work, or (b) laying out the solution in the input" (e.g., Lachter & Bever, 1988). Part of the fascination of connectionist modelling lies in the fact that it offers the experimenter so many surprises. These are self-organizing systems that learn how to solve a problem. As the art is currently practiced, NO ONE fiddles with the internal weights but the system itself, in the course of learning. Indeed, in a simulation of any interesting level of complexity, it would be virtually impossible to reach a solution by "hand-tweaking" of the weights. As for the issue of "sneaking the solution into the input", we have seen several simulations in which the Experimenter did indeed try to make the input as explicit as possible — and yet the system stubbornly found a different way to solve the problem. Good connectionist modelers approach their simulations with the same spirit of discovery and breathless anticipation that is very familiar to those who carry out real experiments with real children. Aside from being close to impossible, cheating would not be any fun at all — and the handcrafting of solutions is usually considered a form of cheating.

Worry #4 . "The supposed commitment to neural plausibility is a scam; no one really takes it seriously." Connectionists work at many different levels between brain and behavior. In current simulations of higher cognitive processes, it is true that the architecture is "brain- like" only in a very indirect sense. In fact, the typical 100-neuron connectionist toy is "brain-like" only in comparison with the serial digital computer (which is wildly unlike nervous systems of any known kind). The many qualities that separate real brains from connectionist simulations have been described in detail elsewhere (Hertz, Krogh and Palmer, 1991; Crick, 1989; Churchland and Sejnowski, in press). The real questions are: (a) is there anything of interest that can be learned from simulations in simplified systems, and (b) can connectionists "add in" constraints from real neural systems in a series of systematic steps, approaching something like a realistic theory of mind and brain? Of course we still do not know the answer to either of these questions, but there are many

in the weighted connections between units. Because these networks are non-linear systems, they can behave in unexpected ways, mimicking the U-shaped learning functions and sudden moments of "insight" that challenged old Stimulus-Response theories of learning, and helped to bring about the cognitive revolution in the 1960's (Plunkett & Marchman, 1991a, 1991b; MacWhinney, 1991).

(6) *Emergent form* . Because connectionist networks are non-linear systems, capable of unexpected forms of change, they are also capable of producing truly novel outputs. In trying to achieve stability across a large number of superimposed, distributed patterns, the network may hit on a solution that was "hidden" in bits and pieces of the data; that solution may be transformed and generalized across the system as a whole, resulting in what must be viewed as a qualitative shift. This is the first precise, formal embodiment of the notion of emergent form — an idea that stood at the heart of Piaget's theory of change in cognitive systems. As such, connectionist systems may have the very property that we need to free ourselves from the Nature-Nurture controversy. New structures can emerge at the interface between "nature" (the initial architecture of the system) and "nurture" (the input to which that system is exposed). These new structures are not the result of black magic, or vital forces. They are the result of laws that govern the integration of information in non-linear systems — which brings us to our final section.

(4) Some Common Misconceptions about Connectionism

It is no doubt quite clear to the reader that we are enthusiastic about the Second Computer Metaphor, because we believe that it will help us to pick up a cold trail that Piaget first pioneered, moving toward a truly interactive theory of change. But we are aware of how much there is to do, and how many pitfalls lie before us. We are also aware of some of the doubts and worries about this movement that are currently in circulation. Perhaps it would be useful to end this essay with some answers to some common misconceptions about connectionism, with special reference to the application of connectionist principles within developmental psychology.

Worry #1 . "Connectionism is nothing but associationism, and we already know the limits of associationism" (e.g., Fodor and Pylyshyn, 1988). As we pointed out above, multi-layer connectionist nets are non-linear dynamical systems, whereas the familiar associationist models of the past rested on assumptions of linearity. This is both the good news, and the bad news. The good news is that non-linear systems can learn relationships of considerable complexity, and they can produce surprising and (of course) non-linear forms of change. The bad news is that no one really understands the limits and capabilities of non-linear dynamical systems. Maybe this is also good news: we have finally met our goal, after years of physics envy, because we have finally reached the same frontiers of ignorance as the physicists! Presumably, the limits of these systems will someday be known (although probably not within our lifetimes). But right now, it would be grossly premature to claim that connectionist networks can "never" perform certain functions. Anyone who claims that we al-

a function of learning, displaying forms of organization that were not placed there by the programmer (or by Nature, or by the Hand of God). To be sure, the final product is co-determined by the initial structure of the system and the data to which it is exposed. These systems are not anarchists, nor solipsists. But in no sense is the final product "copied" or programmed in. Furthermore, once the system has learned it is difficult for it to "unlearn", if by "unlearning" we mean a return to its pristine prelearning state. This is true for the reasons described in (1) and (2): the knowledge contained in connectionist nets is contained in and defined by its very architecture, in the connection weights that currently hold among all units as a function of prior learning. Knowledge is not "retrieved" from some passive store, nor is it "placed in" or "passed between" spatially localized buffers. Learning is structural change, and experience involves the activation of potential states in that system as it is currently structured.

From this point of view, the term "acquisition" is an infelicitous way of talking about learning or change. Certain states become possible in the system, but they are not acquired in the usual sense, i.e., found or purchased or stored away like nuts in preparation for the winter. This property of connectionist systems permits us to do away with problems that have been rampant in certain areas of developmental psychology, e.g., the problem of determining "when" a given piece of knowledge is acquired, or "when" a rule finally becomes productive. Instead, development (like the representations and mappings on which it is based) can be viewed as a gradual process; there is no single moment at which learning can be said to occur (but see non-linearity, below).

(4) *Software as Hardware.* We have stated that knowledge in connectionist nets is defined by the very structure of the system. For this reason, the hardware/software distinction is impossible to maintain under the Second Computer Metaphor. This is true whether or not the structure of connectionist nets as currently conceived is "neurally real", i.e., like the structure that holds in real neural systems. We may still have the details wrong (indeed, we probably do), but the important point for present purposes is that there is no further excuse for ignoring potential neural constraints on proposed cognitive architectures. The distinction that has separated cognitive science and neuroscience for so long has fallen, like the Berlin Wall. Some cognitive psychologists and philosophers of science believe that is not a good thing (and indeed, the same might be said someday for the Berlin Wall). But we are convinced that this historic is a good one, especially for those of us who are interested in the codevelopment of mind and brain. We are going the right direction, even though we have a long way to go.

(5) *Non-linear dynamics* . Connectionist networks are non-linear dynamical systems, a fact that follows from several properties of connectionist architecture including the existence of intervening layers between inputs and outputs (permitting the system to go beyond linear mappings), the non-linear threshold functions that determine how and when a single unit will fire, and the learning rules that bring about a change

graded patterns of breakdown that are typically displayed by brain-damaged individuals (Hinton and Shallice, 1991; Marchman, 1992; Seidenberg & McClelland, 1989; Schwartz, Saffron, & Dell, 1990). Second, the same units can participate in many different patterns, and many different patterns coexist in a super-imposed fashion across the same set of units. This fact that can be used to account for degrees of similarity between patterns, and for the ways in which patterns penetrate, facilitate and/or interfere with one another at various points in learning and development (for an expanded discussion of this point, see Bates, Thal and Marchman, 1991).

(2) *Graded Rules.* Contrary to rumor, it is not the case that connectionist systems have no rules. However, the rules or "mappings" employed by connectionist nets take a very different form from the crisp algorithms contained within the programs employed by a serial digital computer. These include the learning rule itself (i.e., the principle by which the system reduces error and "decides" when it has reached a good fit between input and output), and the functions that determine when and how a unit will fire. But above all, the "rules" in a connectionist net include the connections that hold among units, i.e., the links or "weights" that embody all the potential mappings from input to output across the system as a whole. This means that rules (like representations) can exist by degree, and vary in strength.

It should also be clear from this description that it is difficult to distinguish between rules and representations in a connectionist net. The knowledge or "mapping potential" of a network is comprised of the units that participate in distributed patterns, and the connections among those units. Because all these potential mappings coexist across the same "territory", they must compete with one another to resolve a given input. In the course of this competition, the system does not "decide" between alternatives in the usual sense; rather, it "relaxes" or "resolves" into a (temporary) state of equilibrium. In a stochastic system of this kind, it is possible for several different networks to reach the same solution to a problem, each with a totally different set of weights. This fact runs directly counter to the tendency in traditional cognitive and linguistic research to seek "the rule" or "the grammar" that underlies a set of behavioral regularities. In other words, rules are not absolute in any sense — they can vary by degree within a given individual, and they can also vary in their internal structure from one individual to another. We believe that these properties are far more compatible with the combination of universal tendencies and individual variation that we see in the course of human development, and they are compatible with the remarkable neural and behavioral plasticity that is evident in children who have suffered early brain injury (Thal et al. 1991; Marchman, 1992).

(3) *Learning as structural change.* As we pointed out earlier, much of the current excitement about connectionist systems revolves around their capacity for learning and self-organization. Indeed, the current boom in connectionism has brought learning and development back onto center stage in cognitive science. These systems really do change as

Longuet-Higgins, 1969), and their resemblance to real neural systems continued to exert some appeal (Grossberg, 1968, 1972, 1987). But the current "boom" in parallel distributed processing or "connectionism" was inspired in large measure by the discovery of a learning rule that worked for multi-layered systems (Rumelhart, Hinton and Williams, 1986; Le Cun, 1985). The Minsky-Papert conjecture was overturned, and there are now many impressive demonstrations of learning in multilayered neural nets, including learning of n-order dependencies like "A or B but not both" (Rumelhart and McClelland, 1986 See Footnote 2). Multilayer networks have been shown to be universal function approximators, which means they can approximate any function to any arbitrary degree of precision (Hornik, Stinchcombe, & White; 1989). Such a network is shown in Figure 2.

Another reason for the current popularity of connectionism derives from technical advances in the design of parallel computing systems. It has become increasingly clear to computer scientists that we are close to the absolute physical limits on speed and efficiency in serial systems — and yet the largest and fastest serial computers still cannot come close to the speed with which our small, slow, energy-efficient brains recognize patterns and decide where and how to move. As Carver Mead has pointed out (Mead, 1989), it is time to "reverse-engineer Nature", to figure out the principles by which real brains compute information. It is still the case that most connectionist simulations are actually carried out on serial digital computers (which mimic parallelism by carrying out a set of would-be parallel computations in a series, and waiting for the result until the next wave of would-be parallel computations is ready to go). But new, truly parallel architectures are coming on line (e.g., the now-famous Connection Machine) to implement those discoveries that have been made with pseudo-parallel simulations. Parallel distributed processing appears to be the solution elected by Evolution, and (if Mead is right) computer science will have to move in this direction to capture the kinds of processing that human beings do so well.

For developmental psychologists, the Second Computer Metaphor holds some clear advantages for the study of change in human beings. The first set involves the same four areas in which the First Computer Metaphor has let us down: the nature of representations, rules or "mappings", learning, and the hardware/software issue. The last two are advantages peculiar to connectionist networks: non-linear dynamics, and emergent form.

(1) *Distributed representations* . The representations employed in connectionist nets differ radically from the symbols manipulated by serial digital computers. First, these representations are "coarse-coded", distributed across many different units. Because of this property, it is reasonable to talk about the degree to which a representation is active or the amount of a representation that is currently available in this system (i.e., 50% of an "A" or 99% of the number "7"). This also means that patterns can be built up or torn down in bits and pieces, accounting for the graded nature of learning in most instances, and for the gradual or

input for evidence of its preferred feature; depending on its degree of certainty that the relevant feature has appeared, each demon "shouts" or "whispers" its results. In the Pandemonium Model (as in the Perceptron), there is no final arbiter, no *homunculus* or central executive who puts all these daemonical inputs together. Rather, the "solution" is an emergent property of the system as a whole, a global pattern produced by independent, local computations. This also means that results or solutions can vary in their degree of resemblance to the "right" answer, capturing the rather fuzzy properties of human categorization that are so elusive in psychological models inspired by the serial digital computer.

So far so good. And yet this promising line of research came to a virtual end in 1969, when Minsky and Papert published their famous book *Perceptrons*. Minsky and Papert (who were initial enthusiasts and pioneers in perceptron research) were able to prove that perceptrons are only capable of learning a limited class of first-order, linearly separable patterns. These systems are incapable of learning second-order relations like "A or B but not both" (i.e., logical exclusive OR), and by extension, any pattern of equivalent or greater complexity and inter-dependence. This fatal flaw is a direct product of the fact that perceptrons are two-layered systems, with a single direct link between each input and output unit. If A and B are both "on" in the input layer, then they each automatically "turn on" their collaborators on the output layer. There is simply no place in the system to record the fact that A and B are both on simultaneously, and hence no way to "warn" their various collaborators that they should shut up on this particular trial. It was clear even in 1969 that this problem could be addressed by adding another layer somewhere in the middle, a set of units capable of recording the fact that A and B are both on simultaneously, and therefore capable of inhibiting output nodes that would normally turn on in the presence of either A or B. So why not add a set of "in between" units, creating 3 or 4 or N-layered perceptrons? Unfortunately, the learning rules available at that time (e.g., the simple delta rule) did not work with multilayered systems. Furthermore, Minsky and Papert offered the conjecture that such a learning rule would prove impossible in principle, due to the combinatorial complexity of delta calculations and "distribution of blame" in an n-layered system. As it turns out, this conjecture was wrong (after all, a conjecture is not a proof). Nevertheless, it was very influential. Interest in the perceptron as a model of complex mental processes dwindled in many quarters. From 1970 on, most of artificial intelligence research abandoned this architecture in favor of the fast, flexible and highly programmable serial digital computer. And most of cognitive psychology followed suit. (For a somewhat different account of this history, see Papert, 1988; a good collection of historically important documents can be found in Anderson & Rosenfeld, 1989.).

Parallel distributed processing was revived in the late 1970's and early 1980's, for a variety of reasons. In fact, the computational advantages of such systems were never entirely forgotten (Anderson, 1972; Feldman & Ballard, 1980; Hinton & Anderson, 1981; Kohonen, 1977; Willshaw, Buneman, &

degree or strength, from 0 to 1 (in a purely excitatory system) or from -1 to +1 (in a system with both activation and inhibition). A given output unit would "fire" as a function of the amount of input that it received from the various input units, with activation collected until a critical firing threshold was reached (see also McCulloch and Pitts, 1943). Individual acts of recognition or categorization in a Perceptron reflect the collective activity of all these units. Knowledge is a property of the connection strengths that hold between the respective input and output layers; the machine can be said to "know" a pattern when it gives the correct output for a given class of inputs (including novel members of the input class that it has never seen before, i.e., generalization).

There are some obvious analogies between this system and the form of computation carried out in real neural systems, e.g., excitatory and inhibitory links, summation of activation, firing thresholds, and above all the distribution of patterns across a large number of inter-connected units. But this was not the only advantage that perceptrons offered, compared with their competitors. The most important property of perceptrons was (and is) their ability to learn by example.

During the teaching and learning phase, a stimulus is registered on the input layer in a distributed fashion, by turning units on or off to varying degrees. The system produces the output that it currently prefers (based, in the most extreme *tabula rasa* case, on a random set of connections). Each unit in this distributed but "ignorant" output is then compared with the corresponding unit in the "correct" output. If a given output unit within a distributed pattern has "the right answer", its connection strengths are left unchanged. If a given output has "the wrong answer", the size of the error is calculated by a simple difference score (i.e., "delta"). All of the connections to that erroneous output are then increased or decreased in proportion to the amount of error that they were responsible for on that trial. This procedure then continues in a similar fashion for other trials. Because the network is required to find a single set of connection weights which allow it to respond correctly to all of the patterns it has seen, it typically succeeds only by discovering the underlying generalizations which relate inputs to outputs. The important and interesting result is that the network is then able to respond appropriately not only to stimuli it has seen before, but to novel stimuli as well. The learning procedure is thus an example of learning inductively.

Compared with the cumbersome hypothesis-testing procedures that constitute learning in serial digital computers, learning really appears to be a natural property of the perceptron. Indeed, perceptrons are able to master a broad range of patterns, with realistic generalization to new inputs as a function of their similarity to the initial learning set. The initial success of these artificial systems had some impact on theories of pattern recognition in humans. The most noteworthy example is Selfridge's "Pandemonium Model" (Selfridge, 1958), in which simple local feature detectors or "demons" work in parallel to recognize a complex pattern. Each demon scans the

thereby reducing the number and range of constraints to which our cognitive theories must respond. As a by-product (since divorces usually affect both parties), this approach has also reduced the impact of cognitive theories and cognitive phenomena on the field of neuroscience.

The separation between biology and cognition has had particularly serious consequences for developmental psychology, a field in which biology has traditionally played a major role (i.e., a tradition that includes Freud, Gesell, Baldwin, and Piaget, to name a few). Not only have we turned away from our traditional emphasis on change, but we have also turned away from the healthy and regular use of biological constraints on the study of developing minds. Ironically, some of the strongest claims about innateness in the current literature have been put forth in complete disregard of biological facts. Very rich forms of object perception and deep inferences about three-dimensional space are ascribed to infants before 3 - 4 months of age, conclusions which are difficult to square with (for example) well-known limitations on visual acuity and/or the immaturity of higher cortical regions in that age range. The underlying assumption appears to be that our cognitive findings have priority, and if there is a mismatch between cognitive and biological conclusions, we probably got the biology wrong (which may be the case some of the time — but surely not all the time!).

It seems to us that we need all the constraints that can be found to make sense of a growing mass of information about cognitive development, language development, perceptual development, social development. Furthermore, we suspect that developmental neuroscience would also profit from a healthy dose of knowledge about the behavioral functions of the neural systems under study. Finally, we would all be better off if we could find a computational model (or class of models) in which it would be easier to organize and study the mutual constraints that hold between mental and neural development — which brings us to the next computer metaphor.

(3) The Second Computer Metaphor and Its Implications for Development

During the 1950's and 60's, when the First Computer Metaphor for mind began to influence psychological research, some information scientists were exploring the properties of a different and competing computational device called the Perceptron (Rosenblatt, 1958, 1962). The roots of this approach can be traced to earlier work in cybernetics (Minsky, 1956; von Neumann, 1951, 1958) and in neurophysiology (Eccles, 1953; Hebb, 1949; McCulloch & Pitts, 1943). In a perceptron network, unlike the serial digital computer, there was not a clear distinction between processor and memory, nor did it operate on symbols in the usual sense of the term. Instead, the perceptron network was composed of a large number of relatively simple "local" units that worked in parallel to perceive, recognize and/or categorize an input. These local units or "nodes" were organized into two layers, an "input set" and an "output set". In the typical perceptron architecture, every unit on the input layer was connected by a single link to each and every unit on the output layer (see Figure 1). These connections varied in

Once again, however, we have an unfortunate mismatch between theory and data in cognitive science. Because the hypotheses tested by a traditional computer learning model are discrete in nature (based on the rule and representations described above), learning (a.k.a. "selection") necessarily involves a series of discrete decisions about the truth or falsity of each hypothesis. Hence we would expect change to take place in a crisp, step-wise fashion, as decisions are made, hypotheses are discarded, and new ones are put in their place. But human learning rarely proceeds in this fashion, characterized more often by error, vacillation and backsliding. In fact, the limited value of the serial digital computer as a metaphor for learning is well known. Perhaps for this reason, learning and development have receded into the background in modern cognitive psychology, while the field has concentrated instead on issues like the nature of representation, processes of recognition and retrieval, and the various stages through which discrete bits of information are processed (e.g., various buffers and checkpoints in a serial process of symbol manipulation). Developmental psychologists working within this framework (or indirectly influenced by it) have moved away from the study of change and self-organization toward a catalogue of those representations that are there at the beginning (e.g., the "competent infant" movement in cognition and perception; the parameter-setting movement in developmental psycholinguistics), and/or a characterization of how the processes that elaborate information mature or expand across the childhood years (i.e., changes in performance that "release" the expression of pre-existing knowledge).

(4) *The hardware/software distinction* . One of the most unfortunate consequences of the First Computer Metaphor for cognitive science in general and developmental psychology in particular has been the acceptance of a strong separation between software (the knowledge — symbols, rules, hypotheses, etc. — that is contained in a program) and hardware (the machine that is used to implement that program). From this perspective, the machine itself places very few constraints on our theory of knowledge and (by extension) behavior, except perhaps for some relatively banal concerns about capacity (e.g., there are some programs that one simply cannot run on a small personal computer with limited memory).

The distinction between hardware and software has provided much of the ammunition for an approach to philosophy of mind and cognitive science called Functionalism (Fodor, 1981; See Footnote 1). Within the functionalist school, the essential properties of mind are derived entirely from the domains on which the mind must operate: language, logic, mathematics, three-dimensional space, etc. To be sure, these properties have to be implemented in a machine of some kind, but the machine itself does not place interesting constraints on mental representations (i.e., the objects manipulated by the mind) or functional architecture (i.e., the abstract system that manipulates those objects). This belief has justified an approach to cognition that is entirely independent of neuroscience,

course been known for some time. To resolve this well-known discrepancy, the usual device adopted by proponents of the First Computer Metaphor for Mind is the competence/performance distinction. That is, it is argued that our knowledge (competence) takes a discrete and idealized form that is compatible with the computer metaphor, but our behavior (performance) is degraded by processing factors and other sources of noise that are irrelevant to a characterization of knowledge and (by extension) acquisition of knowledge. This is a perfectly reasonable intellectual move, but as we will see in more detail below, it has led to certain difficulties in characterizing the nature of learning that often result in the statement that learning is impossible.

(2) *Absolute rules.* Like the symbolic representations described above, the algorithms contained in a computer program also take a discrete form. If the discrete symbols that trigger a given rule are present in the input, then that rule must apply, and give an equally discrete symbol or string of symbols as its output. Conversely, if the relevant symbols are not present in the input, then the rule in question will not apply. There is no room for anything in between, no coherent way of talking about 50% of a rule, or (for that matter) weak vs. strong rules. Indeed, this is exactly the reason why computers are so much more reliable than human beings for many computational purposes.

Presented with the well-known mismatch between human behavior and the absolute status of rules in a serial digital computer, proponents of the First Computer Metaphor for Mind usually resort to the same competence/performance described above. Alternatively, there have been attempts to model the probabilistic nature of human behavior by adding weights to rules, a device that permits the model to decide which rule to apply (or in what order of preference) when a choice has to be made. The problem is that these weights are in no way a natural product or property of the architecture in which they are embedded, nor are they produced automatically by the learning process. Instead, these weights are arbitrary, ad hoc devices that must be placed in the system by hand — which brings us to the next point.

(3) *Learning as programming.* The serial digital computer is not a self-organizing system. It does not learn easily. Indeed, the easiest metaphor for learning in a system of this kind is programming; that is, the rules that must be applied to inputs of some kind are placed directly into the system — by man, by Nature or by the hand of God. To be sure, there is a literature on computer learning in the field of artificial intelligence. However, most of these efforts are based on a process of hypothesis testing. In such learning models, two essential factors are provided *a priori* : a set of hypotheses that will be tested against the data, and an algorithm for deciding which hypothesis provides the best fit to those data. This is by its very nature a strong nativist approach to learning. It is not surprising that learning theories of this kind are regularly invoked by linguists and psycholinguists with a strong nativist orientation. There is no graceful way for the system to derive new hypotheses (as opposed to modifications of a pre-existing option). Everything that really counts is already there at the beginning.

deny that society has an influence on development, because we are quite sure that it does. Our point is, simply, that the pendulum has swung too far from the study of child-initiated change. The most influential movements in developmental psychology for the last two decades are those that have deemphasized change in favor of an emphasis on some kind of preformation: either a preformation by Nature and the hand of God, or a preformation by the competent adult.

Why have we accepted these limits? Why haven't we moved on to study the process by which new structures really do emerge? We believe that developmental psychology has been influenced for many years by a metaphor for mind in which it is difficult to think about change in any interesting form — which brings us to the First Computer Metaphor.

(2) The First Computer Metaphor and its Implications for Development

At its core, the serial digital computer is a machine that manipulates symbols. It takes individual symbols (or strings of symbols) as its input, applies a set of stored algorithms (a program) to that input, and produces more symbols (or strings of symbols) as its output. These steps are performed one at a time (albeit very quickly) by a central processor. Because of this serial constraint, problems to be solved by the First Computer must be broken down into a hierarchical structure that permits the machine to reach solutions with maximum efficiency (e.g., moving down a decision tree until a particular subproblem is solved, and then back up again to the next step in the program).

Without question, exploitation of this machine has led to huge advances in virtually every area of science, industry and education. After all, computers can do things that human beings simply cannot do, permitting quantitative advances in information processing and numerical analysis that were unthinkable a century ago. The problem with this device for our purposes here lies not in its utility as a scientific tool, but in its utility as a scientific metaphor, in particular as a metaphor for the human mind/brain. Four properties of the serial digital computer have had particularly unfortunate consequences for the way that we have come to think about mental and neural development.

(1) *Discrete representations.* The symbols that are manipulated by a serial digital computer are discrete entities. That is, they either are or are not present in the input. There is no such thing as 50% of the letter **A** or 99% of the number **7**. For example, if a would-be user types in a password that is off by only one keystroke, the computer does not respond with "What the heck, that's close enough." Instead, the user is damned just as thoroughly as he would be if he did not know the password at all.

People (particularly children) rarely behave like this. We can respond to partial information (degraded input) in a systematic way; and we often transform our inputs (systematic or not) into partial decisions and imperfect acts (degraded output). We are error-prone, but we are also forgiving, flexible, willing and able to make the best of what we have. This mismatch between human behavior and the representations manipulated by serial digital computers has of

career seeking a set of formalisms to concretize his deep insights about change. Most critics agree that these efforts failed. This failure, coupled with new empirical information showing that many other aspects of the theory were incorrect, has led to a widespread repudiation of Piaget. Indeed, we are in a period of "anti-Piagetianism" of patricidal dimensions.

But what have we put in Piaget's place? We have never replaced his theory with a better account of the epistemology of change. In fact, the most influential developmental movements of the last two decades have essentially disavowed change. Alas, we fear that we are back on the horns of the Nature-Nurture dilemma from which Piaget tried in vain to save us.

On the one hand, we have seen a series of strong nativist proposals in the last few years, including proposals by some neo-Gibsonian theorists within the so-called "competent infant movement" (Baillargeon & deVos, 1991; Spelke, 1990, 1991), and proposals within language acquisition inspired by Chomsky's approach to the nature and origins of grammar (Hyams, 1986; Roeper & Williams, 1987; Lightfoot, 1991). In both these movements, it is assumed that the essence of what it means to be human is genetically predetermined. Change — insofar as we see change at all — is attributed to the maturation of predetermined mental content, to the release of preformed material by an environmental "trigger", and/or to the gradual removal of banal sensory and motor limitations that hid all this complex innate knowledge from view. Indeed, the term "learning" has taken on such negative connotations in some quarters that efforts are underway to eliminate it altogether. The following quotes from Piatelli-Palmerini (1989) illustrate how far things have gone:

> I, for one, see no advantage in the preservation of the term learning. We agree with those who maintain that we would gain in clarity if the scientific use of the term were simply discontinued. *(p. 2)*

> Problem-solving...adaptation, simplicity, compensation, equilibration, minimal disturbance and all those universal, parsimony-driven forces of which the natural sciences are so fond, recede into the background. They are either scaled down, at the physico-chemical level, where they still make a lot of sense, or dismissed altogether. *(pp. 13-14)*

On the other hand, the neo-Vygotskian movement and associated approaches to the social bases of cognition have provided us with another form of preformationalism, insisting that the essence of what it means to be human is laid out for the child in the structure of social interactions (Bruner & Sherwood, 1976; Rogoff, 1990). In these theories, change is viewed primarily as a process of internalization, as the child takes in preformed solutions to problems that lie in the "zone of proximal development", i.e., in joint activities that are just outside his current ability to act alone. Related ideas are often found in research on "motherese", i.e., on the special, simplified and caricatured form of language that adults direct to small children (for a review, see Ferguson & Snow, 1978). In citing these examples, we do not want to

tracking, and/or grasping), the child begins to act upon the world (assimilation). Actions are modified in response to feedback from that world (accommodation), and in response to the degree of internal coherence or stability that action schemes bear to one another (reciprocal assimilation). The proximal cause that brings about adaptation is a rather poorly defined notion of equilibration, i.e., the re-establishment of a stable and coherent state after a perturbation that created instability or disequilibrium. In the infant years, adaptation of simple sensorimotor schemes to a structured world leads to an increasingly complex and integrated set of schemes or "plans", structures that eventually permit the child to "re-present" the world (i.e., to call potential perceptuo-motor schemes associated with a given object or event into an organized state-of-readiness, in the absence of direct perceptual input from the represented object or event). This developmental notion of representation comprised Piaget's explanation for the appearance of mental imagery, language and other symbolic or representational forms somewhere in the second year of life. After this point, the process of adaptation continues at both the physical and representation level (i.e., operations on the real world, and operations on the new "mental world"), passing through a series of semi-stable "stages" or moments of system-wide equilibrium, ultimately leading to our human capacity for higher forms of logical and reasoning.

This "bootstrapping" approach to cognitive development does involve a weak form of learning (as defined above), but the mental structures that characterize each stage of development are not predictable in any direct way from either the structure of the world or the set of innate sensorimotor schemes with which the child began. Furthermore, Piaget insisted that these progressive increases in complexity were a result of activity ("construction"), and not a gradual unfolding of predetermined forms (maturation). In this fashion, Piaget strove to save us from the Nature/Nurture dilemma. Behavioral outcomes were determined not only by genes, or by environment, but by the mathematical, physical and biological laws that determine the kinds of solutions that are possible for any given problem. As Piaget once stated in a criticism of his American colleague Noam Chomsky, "That which is inevitable does not have to be innate" (Piaget, 1970a).

There was a period in the history of developmental psychology in which Piagetian theory assumed a degree of orthodoxy that many found stifling. Decades later, it now appears that much of Piaget's theory was wrong in detail. For one thing, it is now clear that the infant's initial stock of innate sensorimotor schemes is far richer than Piaget believed. It is also clear that Piaget overestimated the degree of cross-domain stability that children are likely to display at any given point in development (i.e., the notion of a coherent "stage"). Once the details of his stage theory were proven inadequate, all that really remained were the principles of change that formed the bedrock of Piaget's genetic epistemology — notions of adaptation and equilibration that struck many of his critics as hopelessly vague, and a notion of emergent form that many found downright mystical. Piaget was aware of these problems, and spent the latter part of his

(a) *Maturation* . As the term is typically used in the psychological literature (although this use may not be entirely accurate from a biological perspective), "maturation" refers to the timely appearance or unfolding of behaviors that are predetermined, in their structure and their sequence, by a well-defined genetic program. The role of experience in a strong maturational theory is limited to a "triggering" function (providing the general or specific conditions that allow some predetermined structures to emerge) or a "blocking" function (providing conditions that inhibit the expression of some predetermined event). The environment does not, in and of itself, provide or cause behavioral structure.

(b) *Learning* . "Learning" is typically defined as a systematic change in behavior as a result of experience. Under some interpretations, learning refers to a copying or transfer of structure from the environment to the organism (as in "acquisition" or "internalization"). Under a somewhat weaker interpretation, learning may refer to a shaping or alteration of behavior that is caused by experience, although the resulting behavior does not resemble structures in the environment in any direct or interesting way.

(c) *Development* . As defined by Werner (1948) in his elaboration of the "orthogenetic principle", "development" refers to any positive change in the internal structure of a system, where "positive" is further defined as an increase in the number of internal parts (i.e., differentiation), accompanied by an increase in the amount of organization that holds among those parts. Under this definition, the term "development" is neutral to the genetic or experiential sources of change, and may include emergent forms that are not directly predictable from genes or experience considered separately (i.e., the sum is greater than and qualitatively different from the parts).

Although all three terms have been used to describe behavioral change in the psychological literature, the most difficult and (in our view) most interesting proposals are the ones that have involved emergent form, i.e., changes that are only indirectly related to structure in the genes or the environment. We are referring here not to the banal interactionism in which black and white yield grey, but to a much more challenging interactionism in which black and white converge and interact to yield an unexpected red. Because this interactionist view appears to be the only way to explain how new structures arise, it may be our only way out of a fruitless Nature/Nurture debate that has hampered progress in developmental psychology for most of its history.

Within our field, the most complete interactionist theory of behavioral change to date is the theory offered by Jean Piaget, across a career that spanned more than fifty years (Piaget, 1952, 1970a, 1970b, 1971). Piaget's genetic epistemology concentrated on the way that new mental structures emerge at the interface between an active child and a structured world. The key mechanism for change in Piaget's theory is the consummate biological notion of adaptation. Starting with a restricted set of sensorimotor schemes (i.e., structured "packages" of perception and action that permit activities like sucking, reaching,

CONNECTIONISM AND THE STUDY OF CHANGE

Elizabeth A. Bates & Jeffrey L. Elman

Center for Research in Language
University of California, San Diego

Developmental psychology and developmental neuropsychology have traditionally focussed on the study of children. But these two fields are also supposed to be about the study of change, i.e., changes in behavior, changes in the neural structures that underlie behavior, and changes in the relationship between mind and brain across the course of development. Ironically, there has been relatively little interest in the mechanisms responsible for change in the last 15 - 20 years of developmental research. The reasons for this de-emphasis on change have a great deal to do with a metaphor for mind and brain that has influenced most of experimental psychology, cognitive science and neuropsychology for the last few decades, i.e., the metaphor of the serial digital computer. We will refer to this particular framework for the study of mind as the *First Computer Metaphor*, to be contrasted with a new computer metaphor, variously known as connectionism, parallel distributing processing, and/or neural networks. In this brief chapter, we will argue that the First Computer Metaphor has had some particularly unhappy consequences for the study of mental and neural development. By contrast, the *Second Computer Metaphor* (despite its current and no doubt future limitations) offers some compelling advantages for the study of change, at both the mental and the neural level.

The chapter is organized as follows: (1) a brief discussion of the way that change has (or has not) been treated in the last decade of research in developmental psychology, (2) a discussion of the First Computer Metaphor, and its implications for developmental research, (3) an introduction to the Second Computer Metaphor, and the promise it offers for research on the development of mind and brain, ending with (4) a response to some common misconceptions about connectionism.

(1) What Happened to the Study of Change?

Traditionally, there are three terms that have been used to describe changes in child behavior over time: maturation, learning, and development. For our purposes here, these terms can be defined as follows.

To appear in Mark Johnson (Ed.), *Brain development and cognition: A reader.* Oxford: Blackwell Publishers.

Connectionism and the Study of Change

Elizabeth A. Bates
Jeffrey L. Elman

CRL Technical Report 9202
February 1992

Center for Research in Language

University of California, San Diego
La Jolla, CA 92093-0526