INNATENESS AND EMERGENTISM

Elizabeth Bates[1]
Jeffrey Elman[1]
Mark Johnson[2]
Annette Karmiloff-Smith[2]
Domenico Parisi[3]
Kim Plunkett[4]

[1]***University of California, San Diego***
[2]***MRC Cognitive Development Unit, London***
[3]***National Council of Research, Institute of Psychology, Rome***
[4]***Oxford University***

In William Bechtel & George Graham (Eds.), ***A Companion to Cognitive Science*** **(pp. 590-601). Oxford: Basil Blackwell. Support for the work described here has been provided by NIH/NIDCD-R01-DC00216, NIH-NIDCD P50 DC1289-9351, NIH/NINDS P50 NS22343, and a grant from the John D. and Catherine T. MacArthur Foundation Research Network on Early Childhood Transitions.** **Please address all correspondence to Elizabeth Bates, Center for Research in Language 0526, University of California at San Diego, La Jolla, California 92093-0526, or bates@crl.ucsd.edu.**

INNATENESS AND EMERGENTISM

Elizabeth Bates, Jeffrey Elman, Mark Johnson, Annette Karmiloff-Smith, Domenico Parisi and Kim Plunkett

The Nature-Nurture controversy has been with us since it was first outlined by Plato and Aristotle. Nobody likes it anymore. All reasonable scholars today agree that genes and environment interact to determine complex cognitive outcomes. So why does the controversy persist? First, it persists because it has practical implications that cannot be postponed (i.e., what can we do to avoid bad outcomes and insure better ones?), a state of emergency that sometimes tempts scholars to stake out claims they cannot defend. Second, the controversy persists because we lack a precise, testable theory of the process by which genes and environment interact. In the absence of a better theory, innateness is often confused with (1) *domain specificity* (Outcome X is so peculiar that it must be innate), (2) *species specificity* (we are the only species who do X, so X must lie in the human genome), (3) *localization* (Outcome X is mediated by a particular part of the brain, so X must be innate), and (4) *learnability* (we cannot figure out how X could be learned, so X must be innate). We believe that an explicit and plausible theory of interaction is now around the corner, and that many of the classic maneuvers to defend or attack innateness will soon disappear. In the interim, some serious errors can be avoided if we keep these confounded issues apart. That is the major goal of this paper, i.e., not to attack innateness but to clarify what claims about innateness are (and are not) about.

What will a good theory of interaction look like when it arrives? It is useful here to distinguish between two kinds of interactionism: *simple interactions* (black and white make grey) and *emergent form* (black and white get together and something altogether new and different happens). In an emergentist theory, outcomes can arise for reasons that are not predictable from any of the individual inputs to the problem. Soap bubbles are round because a sphere is the only possible solution to achieving maximum volume with minimum surface (i.e., their spherical form is not explained by the soap, the water, or the little boy who blows the bubble). Beehives take an hexagonal form because that is the stable solution to the problem of packing circles together (i.e., the hexagon is not predictable from the wax, the honey it contains, nor from the packing behavior of an individual bee). D'Arcy Thompson (1917/1968) offered hundreds of examples like these to explain the emergence of different bodily forms, up and down the phylogenetic scale. Jean Piaget argued that logic and knowledge emerge in just such a fashion, from successive interactions between sensorimotor activity and a structured world. In the same vein, it has been argued that grammars represent the class of possible solutions to the problem of mapping hyperdimensional meanings onto a low-dimensional channel, heavily constrained by the limits of human information processing (e.g., MacWhinney & Bates, 1989). Logic, knowledge and grammar are not given in the world, but neither are they given in the genes.

Emergentist solutions of this kind have been proposed again and again in the developmental literature, as a way out of the Nature-Nurture controversy ("That which is inevitable does not have to be innate"). Unfortunately, the metaphors invoked by proponents of emergentism do not constitute a convincing theory of complex cognition, and the detailed descriptions of behavioral change offered by Piagetian scholars have never yielded up the formal theory of development that Piaget sought for more than six decades. As a result, ardent nativists have viewed Piaget as a radical empiricist indistinguishable from Skinner in his reliance on environment as the ultimate cause of development (Chomsky, 1980). A similar fate has befallen those who propose an emergentist account of language (Gibson, 1992).

We believe that a more convincing emergentist account of development is now possible, for three reasons. First, developmentalists have begun to make use of insights from nonlinear dynamics (Elman et al., in press; Thelen & Smith, 1994). This is the latest and perhaps the last frontier of theoretical physics, offering insights into the processes by which complex, surprising and apparently discontinuous outcomes can arise from small quantitative changes along a single dimension. Beehive metaphors have thus given way to an explicit, formal account of emergent form. Second, it is now possible to simulate behavioral change in multilayered neural networks, systems that embody the nonlinear dynamical principles required to explain the emergence of complex solutions from simpler inputs (Elman et al., in press; Rumelhart & McClelland, 1986). Third, students of behavioral development are becoming aware of some remarkable breakthroughs in developmental neurobiology. As we shall see, today's neurobiological results are very bad news for yesterday's nativists, because they underscore the extraordinarily plastic and activity-dependent nature of cortical specialization, and buttress the case for an emergentist

approach to the development of higher cognitive functions.

Even within an interactionist view of this kind, one has to start somewhere. The constraints on emergent form offered by genes and environment must be specified. What do we mean when we say that a given outcome is innately constrained? As a first approximation, we can define "innateness" as a claim about the amount of information in a complex outcome that was contributed by the genes (keeping in mind, of course, that genes do not act independently, and that they can be turned on and off by environmental signals throughout the lifetime of the organism). Elman et al. have proposed a 3-level taxonomy of claims about innateness, ordered from strong to weak with regard to the amount of information that must be contributed by the genes for this claim to work. Each level is operationally defined in terms that correspond to real brains and to artificial neural networks, as follows:

I. Representational constraints refer to direct innate structuring of the mental/neural representations that underlie and constitute "knowledge". Synaptic connectivity at the cortical level is the most likely candidate for the implementation of detailed knowledge in real live brains, because that is the only level that has the coding power for higher-order cognitive outcomes. In artificial neural networks, this level is operationalized in the weighted connections between processing units.

II. Architectural constraints refer to innate structuring of the information-processing system that must acquire and/or contain these representations. Although representation and architecture are not the same thing, there is no question that the range of representations a system can take is strongly constrained at the architectural level. In traditional serial digital computers, some programs can only run on a machine with the right size, speed and power. In neural networks, some forms of knowledge can only be realized or acquired in a system with the right structure (the right number of units, number of layers, types of connectivity between layers, etc.). In fact, there is now a whole subfield of neural network research in which genetic algorithms are applied to uncover the class of architectures that are best suited to a given class of learning problems (Elman, this volume).

To operationalize architectural constraints in real brains and in neural nets, Elman et al. break things down into three sublevels:

A. Basic computing units. In real brains, this sublevel refers to neuronal types, their firing thresholds, neurotransmitters, excitatory/inhibitory properties, etc. In neural networks, it refers to computing elements with their activation function, learning algorithm, temperature, momentum and learning rate, etc.

B. Local architecture. In real brains, this sublevel refers to regional factors like the number and thickness of layers, density of different cell types within layers, type of neural circuitry (e.g., with or without recurrence). In neural networks, it refers to factors like the number of layers, density of units within layers, presence/absence of recurrent feedback units, and so forth.

C. Global architecture. In real brains, this sublevel includes gross architectural facts like the characteristic sources of input (afferent pathways) and patterns of output (efferent pathways) that connect brain regions to the outside world and to one another. In many neural network models, the size of the system is so small that the distinction between local and global architecture is not useful. However, in so-called modular networks or expert networks, it is often useful to talk about distinct subnets and their interconnectivity.

III. Chronotopic constraints refer to innate constraints on the timing of developmental events, including spatio-temporal interactions. In real brains, this would include constraints on the number of cell divisions that take place in neurogenesis, spatio-temporal waves of synaptic growth and pruning, and relative differences in timing between subsystems (e.g., differences among vision, audition, etc. in the timing of thalamic innervation of the cortex). The same level is captured in neural networks by incremental presentation of data, cell division schedules in growing networks, adaptive learning rates, and intrinsic changes in learning that come about because of node saturation.

The reader is referred to Elman et al. for detailed examples at all of these levels. For our purposes here, the point is that strong nativist claims about language (Fodor, 1983; Pinker, 1994), physics (Spelke, 1991) or social reasoning (Horgan, 1995; Leslie, 1994) have to assume representational nativism, implicitly or explicitly, because that is the only level with the required coding power for the implementation of knowledge that is independent of experience. For example, Noam Chomsky (1975) has proposed that "Linguistic theory, the theory of UG [Universal Grammar]... is an innate property of the human mind" (p. 34), and that we should conceive of "the growth of language as analogous to the development of a bodily organ" (p. 11). The mental organ metaphor leaves little room for learning. Indeed, Chomsky has argued that "a

general learning theory....seems to me dubious, unargued, and without any empirical support (1980a, p. 110). Piatelli-Palmarini (1989, p. 2) echoes this theme, stating that "I...see no advantage in the preservation of the term 'learning.' I agree with those who maintain that we would gain in clarity if the scientific use of the term were simply discontinued." Where would such rich innate structure reside? Pinker suggests that this innate knowledge must lie in the "microcircuitry" of the brain. We think that he is absolutely right: If the notion of a language instinct means anything at all, it must refer to a claim about cortical microcircuitry, because this is (to the best of our knowledge) the only way that detailed information can be laid out in the brain.

This kind of representational nativism is theoretically plausible and attractive, but it has proven hard to defend on both mathematical and empirical grounds.

On mathematical grounds, it is difficult to understand how 10^{14} synaptic connections in the human brain could be controlled by a genome with approximately 10^6 genes, particularly when (a) 20-30% of these genes at most go into the construction of a nervous system (Wills, 1991), and (b) humans share approximately 98% of their genes with their nearest primate neighbors (King & Wilson, 1975). But the problem is even worse than that. Paul Churchland (1995) reminds us that each synaptic connection can take multiple values. If we assume conservatively that each connection can take 10 values, Churchland calculates that the synaptic coding power of the human brain contains more potential states of connectivity than there are particles in the universe! Genes would need a lot of information to orchestrate a system of this size. Of course a detailed mapping from genes to cortex would still be possible if genes behaved like letters in the alphabet, yielding up an indefinite set of combinations. But this is not the case; instead, genes operate within a highly constrained spatiotemporal and chemical matrix (Edelman, 1987; Wills, 1991), using and reusing topological principles that have been conserved over millions of years and thousands of species.

One could argue that the innate component of knowledge occupies only a fraction of this massive state space, despite its richness. However, the past two decades of research on vertebrate brain development suggest that fine-grained patterns of cortical connectivity are largely determined by cortical input (for reviews, see Elman et al., chapter 5; Johnson, in press). For example, we know that auditory cortex will take on retinotopic maps if input from the eye is diverted there from its normal visual target (Sur, Pallas, & Roe, 1990), that plugs of cortex taken from one cortical area and transplanted in another will take on the representations that are appropriate for the input they receive in their new home (Stanfield & O'Leary, 1985), that alterations in the body surface of an infant rat lead to corresponding alterations in the cortical map (Killackey, 1990), that the "where is it?" system will take over the "what is it?" function in infant monkeys with bilateral lesions to inferior temporal cortex (Webster, Bachevalier, & Ungerleider, 1995), and that human infants with left-hemisphere lesions that would lead to irreversible aphasia in an adult go on to attain language abilities that are well within the normal range (Bates et al., in press; Eisele & Aram, 1995). In short, there is very little evidence today in support of the idea that genes code for synaptic connectivity at the cortical level. Instead, brain development in higher vertebrates appears to involve massive overproduction of elements early in life (neurons, axons and synapses), followed by a competitive process through which successful elements are kept and those that fail are eliminated (Edelman, 1987). Pasco Rakic refers to this competition as the process by which experience literally "sculpts" the brain. In addition to this sculpting through regression, experience may also add structure across the course of life, inducing synaptic sprouting in just those areas that are challenged by a brand-new task (Greenough, Black, & Wallace, 1993; Merzenich, 1995; Pons et al., 1991).

Although there is surprisingly little evidence for innate representations at the cortical level (cf. Balaban, Teillet, & Le Douarin, 1988), there is substantial evidence for innate architectures and innate variations in timing. This includes evidence that neurons "know" where they are supposed to go during cell migration (Rakic, 1988), and evidence that axons prefer particular targets during their long voyage from one region to another (Niederer, Maimon, & Finlay, 1995; but see Molnar & Blakemore, 1991). Could this kind of innateness provide the basis for an innate Universal Grammar? Probably not, because (a) these gross architectural biases do not contain the coding power required for something as detailed and specific as grammatical knowledge, and (b) the rules of growth at this level appear to operate across species to a remarkable degree. For example, Deacon (in press) describes evidence for lawful axon growth in the brain of the adult rat, from cortical transplants taken from fetal pigs!

There are also regional variations in the neurochemical substrate (e.g., somatosensory cortex transplanted to a visual region will take on visual maps, but still expresses the neurochemicals appropriate for a somatosensory zone—Cohen-Tannoudji, Babinet, & Wassef, 1994), and regional variations in cell density (e.g., primary visual cortex is exceptionally dense, a characteristic that seems to be determined during neurogenesis, before any information is received—Kennedy, Dehay, & Horsburg, 1990). Hence cortical regions are likely to differ from the outset in style of computation, which means that they will also vary in the kinds of tasks they can perform best. In

other words, the competition that characterizes brain development does *not* take place on an even playing field. The game is rigged from the beginning to privilege some overall "brain plans" over others. However, it is also clear that many alternative brain plans are available if the optimal form is precluded for some reason.

Bates et al. (in press) have argued that left-hemisphere specialization for language in humans depends on indirect, architectural constraints like these. The temporal and frontal regions of the left hemisphere play a major role in the mediation of language production in over 95% of normal adults, leading to irreversible aphasia if specific left-hemisphere sites are damaged. And yet, as noted above, infants with homologous injuries do not grow up to be aphasic. How can that be? If left perisylvian cortex isn't necessary for normal language, where does the typical "adult brain plan" come from? Studies of infants with focal brain injury show that the temporal (but not the frontal) region of the left hemisphere is indeed specialized at birth, because children with left temporal lesions are significantly slower in the development of expressive (but not receptive) vocabulary and grammar. However, this regional difference is no longer detectable by the time children with the same early injuries are 7 years old, which means that a great deal of reorganization must have taken place across the first years of life. Evidently other regions of the brain are capable of taking on the representations required for normal language. Bates et al. suggest that left temporal cortex is initially specialized not for language itself, but for the extraction of perceptual detail (e.g., damage to the same regions has specific effects on the extraction of detail from a visual-spatial pattern). Under normal conditions, this indirect bias in computing style leads to left-hemisphere specialization for language. But the representations required for language are not (and apparently need not be) present from the beginning, because the same cat can be skinned in a number of alternative ways.

Because the evidence is not good for strong, representational forms of nativism, the differences that we observe from one species to another must be captured primarily by architectural and chronotopic facts. The final product emerges from the interaction between these constraints and the specific problems that an organism with such structure encounters in the world. Within this framework, let us reconsider some of the classical arguments for innate knowledge—arguments which, we believe, confuse levels of analysis that should be kept separate.

Innateness and domain specificity. It has been argued that language is so peculiar, so specific to the domain in question, that it could not possibly be learned or processed by a domain-general system. Similar claims have been made about face perception, music, mathematics, and social reasoning. Elman et al. argue that claims about domain specificity must (like innateness) be broken down into different levels before we can approach the issue empirically. Using language as our test domain, here is a brief overview.

Behavioral specificity. Languages represent a class of solutions to a problem that is undeniably unique in its scope and nature: the problem of mapping a hyperdimensional meaning space onto a low-dimensional channel (MacWhinney & Bates, 1989). There may be a casual resemblance to domains like birdsong (learning in the vocal channel), chess (a complex set of solutions to a game that only humans play) or music (rule-governed transitions in sound), but these similarities are largely superficial. Turkish case inflections do not "look like" chess, birdsong or music—but they do look a lot like case inflections in Hungarian. That is, languages have very little in common with other cognitive systems, but they do have a lot in common with each other. Where do these commonalities come from? The meaning space involved in the language-mapping problem includes experiences that are shared by all normal members of the species, and the channels used by human language are subject to universal constraints on information processing (e.g., perception, memory, articulatory planning). Under these circumstances, we should not be surprised to find that the class of solutions to the problem is quite limited, constituting a set of alternatives that might be referred to as Universal Grammar. We will stipulate that domain-specific behaviors have emerged in response to this mapping problem, and that natural languages draw from a common set of domain-specific solutions. But such facts do not constitute *ipso facto* evidence for innateness, because the same solutions could have emerged by an emergentist scenario.

Representational specificity. If an individual reliably produces the behaviors required to solve a domain-specific problem, it follows that s/he must possess a set of domain-specific mental/neural representations that support the behavior. That is, every representation must be implemented in a form that is somehow distinguishable from other aspects of knowledge (see localization, below). This generalization holds whether the representations in question are innate or learned; hence the specificity of a representation is simply not relevant to the innateness debate.

Specificity of mental/neural processes. This is the level at which innateness and domain specificity finally cross: Is it possible for a domain-general architecture to acquire and/or process domain-specific representations? Notice that "domain-general" need not mean "a device that can learn and do anything." We have already stipulated the need for a good match between problems and architectures in neural network research. There is no device that can learn and do everything. The debate is more specific: Can a domain

like language be learned and/or processed by *any* system that is not specifically tailored for and dedicated to linguistic events? This is an empirical question, and the answer is not in. However, evidence in support of the domain-general view is available from simulations of domain-specific learning in general-purpose neural networks (see learnability, below), and from the plastic reorganization of language and other higher cognitive functions observed in children with focal brain injury.

But what if the representation at issue is bizarre, and not at all predictable (as far as we can see) from the problem to be solved? How could a general architecture ever acquire such a thing? Our belief that a structure is inexplicable may be nothing more than a comment on our ignorance. For example, the visual cortex of the cat contains odd little neurons that only fire to lines at a particular orientation (Hubel & Wiesel, 1963). Why should such a peculiar structure emerge? And yet we know that such structures do emerge reliably every time a multilayered neural network is forced to extract three-dimensional information from a two-dimensional array (e.g. Miller, 1994; Shatz, 1996). We cannot predict the line-orientation solution just by looking at the inputs to the problem of mapping three dimensions onto two, but apparently just such a solution is required, whether it is built in or not. In other words, although the line orientation detectors in visual cortex *could* be innate (phylogeny insuring a useful solution), these simulations show that they do not *have* to be innate. The same may be true for the odd-looking structures that comprise human grammars, and human social reasoning.

Genetic specificity. Skipping over the intervening levels, arguments linking domain specificity and innateness are sometimes based on the specific patterns of impairment observed in individuals with genetic damage. Suppose, for example, that we uncover a form of language impairment that is associated with a genetically transmitted disorder. Doesn't this constitute direct evidence for the innateness of a domain-specific ability? Not necessarily. After all, language is entirely absent in cases of cerebral agenesis (where no brain grows at all above the brainstem level), but no one would argue that the absence of a brain provides interesting evidence for a domain-specific language faculty. Genetically based language disorders provide evidence for the innateness of a domain-specific faculty if and only if we can show that the genetic defect affects language *in isolation*.

Specific Language Impairment or SLI is defined as a significant disorder in which language falls well below mental age, in the absence of mental retardation, frank neurological impairment, hearing loss, severe social-emotional distress or environmental pathology (see Leonard, in press, for a review). It has been shown that SLI tends to run in families, and some have argued that this disorder constitutes the required evidence for a genetic defect that only affects grammar. A celebrated case in point is the London family in which (it was reported) a genetically transmitted impairment was observed that only affects regular grammatical morphemes (e.g., walk --> walked), with no other effect on any other aspect of cognition or language, including irregular grammatical morphemes (e.g., give --> gave). The initial report generated a great deal of excitement (Gopnik, 1990; Pinker, 1994), but it was ultimately shown to be premature and largely incorrect. More comprehensive studies show that the affected members of this family suffer from a host of deficits inside and outside of language, and the putative dissociation between regular and irregular morphemes does not replicate (Vargha-Khadem, Watkins, Alcock, Fletcher, & Passingham, 1995).

Thirty years of research on other children and adults with SLI yield a similar conclusion (Leonard, 1996): Specific language impairment correlates with a range of relatively subtle deficits outside the boundaries of language proper, including aspects of attention, symbolic play, mental imagery, and the detection of rapid sequences of sounds. In short, Specific Language Impairment is a genetically transmitted disorder, but it is no longer clear (despite the name) that it is specific to language, much less to some peculiarity of grammar.

The converse is also true: Deficits specific to language have turned out not to be innate, at least not in any interesting sense. For example, we now know that grammatical morphology (including all those "little words" and endings) is an especially vulnerable domain; whether or not it is impaired in isolation, morphology shows up as a major problem area in SLI, Down Syndrome, and other populations where a genetic base is known or suspected. However, specific problems with grammatical morphology have also been shown in many different forms of acquired brain injury (with little respect for lesion site), in neurologically intact individuals with hearing impairment, and in college students forced to process sentences under adverse conditions (e.g., with perceptually degraded stimuli, or with reduced attention and memory due to a competing task). Grammatical morphemes tend to be low in perceptual salience and imageability, and perhaps for this reason, they constitute a "weak link in the processing chain." The fact that they are preferentially disrupted in genetically based syndromes does not necessarily constitute evidence that they are innate in any domain-specific way. Damage to the human elbow has a very specific effect on tennis, but that does not mean that the elbow is a tennis processor, nor that the genes that participate in elbow construction do so for the good of tennis. We have already stipulated that language is not tennis, but the metaphor is appropriate on this particular point.

To summarize, innateness and domain specificity are not the same thing, and the case for innateness can never be made simply by listing strange phenomena (i.e., the Madame Tussaud strategy). We turn now to

species specificity, localization and learnability, special cases of the effort to prove innateness by showing that a domain is "special."

Innateness and species specificity. In this variant of the domain specificity approach, it is argued a domain must be innate because only humans do it—or, at least, only humans do it very well. This would include language, but it also includes music, politics, religion, international finance and ice hockey. To be sure, there are rudimentary variants of human skills in other species, including language. The identification of such infra-human precursors is useful, because it can tell us something about the evolution of language and other uniquely human functions. But the case for an innate, domain-specific system cannot be made simply by pointing out that nobody else has what we have. Although we are the only species that plays chess, no one wants to argue that we start out with a chess faculty in any interesting sense.

Of course many of us don't play chess, but all normal humans use language. Do species specificity and universality together constitute evidence for an innate faculty? Possibly, but both facts could be explained by factors that are only indirectly related to the domain in question. To date, no one has ever identified a neural structure that is unique to humans, i.e., a human-specific neuronal type, neurotransmitter, pattern of cortical layering, or even (depending on how we define "area") a human-specific area of the brain (Deacon, in press; Finlay & Darlington, 1995). Our undeniably unique array of skills appears to be built out on quantitative variations in the primate brain plan, e.g., expansion of frontal cortex relative to other regions, proportional enlargement of secondary areas within visual cortex, more direct cortical control over the mouth and fingers. The latter innovation sounds like it might have emerged especially for language (or, more generally, for culture), and at some level that may be the case. It is interesting to note, however, that the same direct connections from cortex to the periphery are present in the embryonic rat, but they are eliminated before they have a chance to become functional (Deacon, in press). Although we do not belong to the school of evolution that explains everything through brain size, species-specific abilities could be an unintended by-product of a much more general change in computing power (Wills, 1991). Species specificity alone does not constitute evidence for a specific mental organ.

Innateness and localization. This is also a form of the specificity argument: Mental organs are special because they take place in their own part of the brain, what Fodor (1983, p. 99) calls a "specialized neural architecture" (a term that conflates the representational and architectural levels laid out by Elman et al.). If we could show, for example, that the brain handles regular and irregular grammatical morphemes differently, wouldn't that constitute evidence for two innately specialized, domain-specific processors? Not necessarily.

First, everything that we know is mediated by the brain. If we experience two stimuli in exactly the same way, then (by definition) we do not "know" that they are different. If we do experience them differently, then that difference must be reflected somewhere in the brain. Every new piece of learning changes the structure of the brain in some fashion, however minor. Consider, for example, a recent demonstration that chess experts show different patterns of cortical activity at different points in the game (Nichelli et al., 1994). This does not mean we have an End Game Organ, not even in the adult state. And it certainly does not mean that we were born with one. All knowledge presupposes localization in some form (compact and local, or broadly distributed), and hence demonstrations of localization do not constitute evidence for innateness. This is true whether the localization is universal (all humans show the same pattern) or variable (some people handle the same content in different places—Caplan, 1981).

However, the converse is not true: If a cognitive ability is innate, then it must be realized in some topographically specifiable way. That's how genes work, i.e., by coding proteins in a spatially, temporally and chemically defined matrix (Edelman, 1987; Wills, 1991). That is precisely why the evidence for cortical plasticity is so devastating to representational nativism. To fend off this evidence, one might envision a scenario in which the genes that set up the nervous system travel around in the bloodstream looking for a friendly environment in which a specific mental organ can be built. After all, every cell in the body contains the entire genome. Perhaps the language genes are wandering about, waiting for a signal that says "Start building a language organ now, here." There are certainly examples in the literature where the right thing does get built in the wrong place, or at least an atypical place (e.g., the master gene for the eye, which can be multiplied in various places). But this kind of evidence appears to be the exception. And at least in that case (in contrast to higher cognitive functions like language), there is a specifiable set of physical constants involved dictating the shape of the thing to be built. For the same reason that we cannot really build a dinosaur out of genes in a piece of amber (the Jurassic Park scenario), genes for language or music or face perception do not travel around in a lifeboat looking for a place to land. Localization does not presuppose innateness, but claims about innateness do presuppose a physical base. That is why nativists are wise to look for the neural correlates of the system they are interested in.

Innateness and learnability. Within linguistics, claims about innateness have been made that bypass all these lines of empirical evidence. The ultimate form of the eccentricity argument goes like this: X (usually language) is so peculiar, so unlike

anything else that we do, that it cannot be learned by garden-variety learning mechanisms (e.g. Crain, 1991). Children (it is claimed) must acquire a grammar that is more powerful than his/her degenerate input can support. They are only able to go beyond their data and zero in on the right grammatical target because they already know a great deal about the class of possible grammars. The most principled form of this argument is based on a formal proof of learnability in computer science called Gold's Theorem (Gold, 1967), which showed that grammars of a particular class cannot be induced or "guessed" from a finite base of positive evidence (i.e., examples of sentences in the language) in the absence of negative evidence (i.e., examples of sentences that do not belong in the language).

A thorough or even a superficial treatment of this argument goes far beyond our purview here, except to note that all learnability proofs rest upon at least four kinds of assumptions: a definition of the grammar to be acquired (e.g., grammars defined as strings of symbols generated by one or more recursive rules), a characterization of the data available to the learner, a specification of the learning device that goes to work on these data, and a criterion that defines successful learning. If the grammar to be acquired is very abstract, if our criterion for success is very high, and/or if the learning device is weak and the data are degenerate, then it follows incontrovertibly that the grammar cannot be learned without a great deal of innate knowledge to make up for those weaknesses. Does any of this apply to human language? As it turns out, Gold's Theorem only applies if we make assumptions about the learning device that are wildly unlike any known nervous system. And that is the only formal proof around at this writing. No one has done the work to find out whether grammars of a different kind are learnable (e.g., probabilistic mappings from meanings onto sound), or whether a learning device with vastly different properties could acquire such a grammar (e.g., a multilayered neural network).

In the interim, there are now simulations of grammatical learning in neural networks that could be viewed as learnability proofs of a sort (for a review, see Elman et al.). For example, Elman (1993) has shown that an artificial grammar with center embeddings and long-distance dependencies (i.e., agreement marking) can be learned by a simple recurrent network that lives in time and tries to guess what is coming next, one word at a time. The system accomplishes this with a positive data base only (i.e., feedback comes in the form of guesses that are confirmed or disconfirmed when the next word comes in). It also makes errors, and then recovers from those errors, in the absence of negative evidence, providing *prima facie* evidence against the generality of Gold's Theorem for language learning in a different kind of system. However, it is not the case that any neural network can accomplish this task. Elman discovered that his network could only learn the grammar if it was first exposed to simple sentences, with complex sentences introduced later. But of course, this is not true for human children, who hear at least a few embedded sentences from the very beginning. Elman found that he could obtain the same result by a simple trick: Start the system out with a rapidly fading memory (instantiated in the units that copy the system's internal state on a previous trial), gradually increasing that memory (independent of learning itself) up to the adult form. As a result, the network could only learn off short strings in the early stages of learning—even though simple and complex strings were both available in the input from the very beginning. This single example illustrates our earlier point about different levels of innateness: A grammar that was unlearnable under one set of timing conditions becomes learnable in a recurrent network when the timing conditions change—all of this accomplished without building innate representations into the architecture before learning begins.

In short, we cannot conclude from the presence of eccentric structures that those structures are innate—not even if they are unique to our species, universal among all normal members of that species, localized in particular parts of the system, and learnable only under specific conditions. The same facts can be explained by replacing innate knowledge (i.e. representations) with architectural and temporal constraints that require much less genetically specified information. This kind of emergentist solution to the Nature-Nurture controversy has been around for many years, but it has only become a scientifically viable alternative in the last decade. As a result, the long-awaited reconciliation between Plato and Aristotle may be at hand.

REFERENCES

Balaban, E., Teillet, M.-A., & N. Le Douarin (1988). Application of the quail-chick chimeric system to the study of brain development and behavior. *Science, 241*, 1339-1342.

Bates, E., Thal, D., Aram, D., Eisele, J., Nass, R., & Trauner, D. (in press). From first words to grammar in children with focal brain injury. To appear in D. Thal & J. Reilly, (Eds.), *Special issue on Origins of Communication Disorders, Developmental Neuropsychology*.

Caplan, D. (1981). On the cerebral organization of linguistic functions: Logical and empirical issues surrounding deficit analysis and functional localization. *Brain and Language, 14*, 120-137.

Chomsky, N. (1975). *Reflections on language*. New York: Parthenon Press.

Chomsky, N. (1980). On cognitive structures and their development: A reply to Piaget. In M. Piatelli-Palmarini (Ed.), *Language and learning: the debate between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.

Churchland, P. M. (1995). *The engine of reason, the seat of the soul: A philosophical journey into the brain.* Cambridge, MA: MIT Press.

Cohen-Tannoudji, M., Babinet, C., & Wassef, M. (1994). Early determination of a mouse somatosensory cortex marker. *Nature, 368(6470*), 460-463.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences, 14*, 597-611.

Deacon, T.W. (in press*). The idea that changed the brain: Coevolution of the human brain and language*. New York: Norton.

Edelman, G.M. (1987). *Neural Darwinism: The theory of neuronal group selection.* New York: Basic Books.

Eisele, J., & Aram, D. (1995). Lexical and grammatical development in children with early hemisphere damage: A cross-sectional view from birth to adolescence. In Paul Fletcher & Brian MacWhinney (Eds.), *The handbook of child language* (pp. 664-689). Oxford: Basil Blackwell.

Elman, J. L. (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. (in press). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press/Bradford Books.

Finlay, B., & Darlington, R. (1995). Linked regularities in the development and evolution of mammalian brains. *Science, 268*, 1578-1584.

Fodor, J.A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MIT Press.

Gibson, E. (1992). On the adequacy of the Competition Model. *Language, 68*, 812-830.

Gold, E.M. (1967). Language identification in the limit. *Information and Control, 16*, 447-474.

Gopnik, M. (1990). Feature-blind grammar and dysphasia. *Nature, 344(6268)*, 715.

Greenough, W.T., Black, J.E., & Wallace, C.S. (1993). Experience and brain development. In M. Johnson (Ed.), *Brain development and cognition: A reader* (pp. 290-322). Oxford: Blackwell.

Horgan, J. (1995). The new Social Darwinists. *Scientific American, 273(4)*, 174-181.

Hubel, D.H., & Wiesel, T.N. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *Journal of Neurophysiology, 26*, 944-1002.

Johnson, M.H. (in press). *Developmental cognitive neuroscience: An introduction.* Oxford: Oxford University Press.

Kennedy, H., Dehay, C., & Horsburg, G. (1990). Striate cortex periodicity. *Nature, 348(6301)*, 494.

Killackey, H.P. (1990). Neocortical expansion: An attempt toward relating phylogeny and ontogeny. *Journal of Cognitive Neuroscience, 2*, 1-17.

King, M., & Wilson, A. (1975). Evolution at two levels in humans and chimpanzees. *Science, 188*, 107-116.

Leonard, L.B. (in press). *Specific language impairment*. Cambridge, MA: MIT Press.

Leslie, A.M. (1994). Pretending and believing - Issues in the theory of Tomm. *Cognition, 50(1-3)*, 211-238.

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

Merzenich, M.M. (1995). Cortical plasticity: Shaped, distributed representations of learned behaviors. In B. Julesz & I. Kovacs (Eds.), *Maturational windows and cortical plasticity in human development: Is there a reason for an optimistic view?* Reading, MA: Addison Wesley.

Miller, K. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns thorugh activity-dependent competition between ON- and OFF-center inputs. *Journal of Neuroscience, 14(1*), 409-441.

Molnar, Z., & Blakemore, C. (1991). Lack of regional specificity for connections formed between thalamus and cortex in coculture. *Nature, 351 (6326)*, 475-477.

Nichelli, P., Grafman, J., Pietrini, P., Alway, D., Carton, J.C., & Miletich, R. (1994). Brain activity in chess playing. *Nature, 369 (6477)*, 191.

Niederer, J., Maimon, G., & Finlay, B. (1995). Failure to reroute or compress thalamocortical projection after prenatal posterior cortex lesions. *Society for Neuroscience Abstracts, 21*.

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

Pinker, S. (1994). *The language instinct: How the mind creates language.* New York: William Morrow.

Pons, T.P., Garraghty, P.E., Ommaya, A.K., Kaas, J.H. Taub, E., & Mishkin M. (1991). Massive cortical reorganization after sensory deafferentation in adult macaques [see comments]. *Science, 252(5014)*, 1857-1860.

Rakic, P. (1988). Specification of cerebral cortical areas. *Science, 241*, 170-176.

Rumelhart, D. & McClelland, J. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* Cambridge, Mass.: MIT Press.

Shatz, C. (1996). The emergence of order in visual system development. *Proceedings of the National Academy of Sciences, 93*, 602-608.

Spelke, E.S. (1991). Physical knowledge in infancy: Reflections on Piaget's theory. In S. Carey & R. Gelman, (Eds.), *Epigenesis of the mind: Essays in biology and knowledge.* Hillsdale, NJ: Erlbaum.

Stanfield, B.B., & O'Leary, D.D. (1985). Fetal occipital cortical neurones transplanted to the rostral cortex can extend and maintain a pyramidal tract axon. *Nature, 313(5998)*, 135-137.

Sur, M., Pallas, S.L., & Roe, A.W. (1990). Cross-modal plasticity in cortical development: Differentiation and specification of sensory neocortex. *TINS, 13*, 227-233.

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

Thompson, D.W. (1917/1968). *On growth and form* (2d ed., reprinted). Cambridge [Eng.] University Press. (Original work published 1917.)

Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences USA, 92*, 930-933.

Webster, M.J., Bachevalier, J., & Ungerleider, L.G. (1995). Development and plasticity of visual memory circuits. In B. Julesz & I. Kovacs (Eds.), *Maturational windows and adult cortical plasticity in human development: Is there reason for an optimistic view?* Reading, MA: Addison-Wesley.

Wills, C. (1991). *Exons, introns, and talking genes: The science behind the Human Genome Project.* New York: Basic Books.