CHAPTER 7

Rethinking innateness

Where does knowledge come from?

This was the question we began with. As we suggested at the outset, the problem is not so much that we do not know what the sources of knowledge are. The problem is rather in knowing how these sources combine and interact. The answer is not Nature *or* Nurture; it's Nature *and* Nurture. But to say that is to trade one platitude for another; what is necessary is to understand the nature of that interaction.

We stated at the beginning that including the word "innate" in our title gave us misgivings. There can be no question about the major role played by our biological inheritance in determining our physical form and our behaviors. We are not empiricists. What troubles us about the term innate is that, as it is often used in the cognitive and developmental sciences, it suggests an overly simplistic view of how development unfolds. To say that a behavior is innate is often taken to mean—in the extreme case—that there is a single genetic locus or set of genes which have the specific function of producing the behavior in question, and only that behavior. To say that a behavior is innate is thus seen as tantamount to explaining the ontogeny of that behavior. In fact, nothing has been explained. And when innate is used more reasonably to refer in general to our genetic endowment, the term ends up being almost vacuous, given the enormous complexity of interactions which are involved in the expression of that endowment.

So why did we use the word at all? In part, it is precisely because the concept of innateness is in such wide use that we wanted to highlight it in the title. There is no avoiding innateness. It has both widespread use in the scientific community, and, unfortu-

nately, figures prominently in the recurring debates about the supposed genetic bases for intelligence. We chose therefore to take the bull by the horns.

But our purpose is more ambitious. We are in fact interested in understanding those phenomena which are often said to be innate. In an important sense we agree that an enormous amount of our behavior *is* innate—only what we mean by the word is quite different than what is often meant. Our goal in this book has been to provide a framework and some tools for understanding the details of the mechanisms by which nature maximizes the likelihood of achieving good solutions to difficult problems. This task is complicated by the fact that all nature cares about is outcomes. Optimal and efficient solutions count for little; tidiness and elegance count for even less.

Despite this, we feel that neuroscientific and modeling techniques have advanced to the point where we can begin to make sense of the developmental process in some concrete detail. We regard this book as a tiny step in this direction. What we propose is not yet a theory, but may hopefully be the beginning of one. It might best be thought of as a framework, a set of conceptual tools for thinking about issues such as development and innateness.

In this final chapter we wish to tie together the various strands laid out in earlier chapters. We begin by summarizing what we see as the primary lessons that can be drawn from the work we have presented. Some of these lessons are negative—warnings about the possible overinterpretation and misinterpretation of phenomena which have been taken as strong indicators of innate bases for behavior. Other lessons are positive—new ways of understanding these same phenomena. Finally, we turn to the future. We recognize that we have but touched the tip of a very large iceberg and much remains to be done. In the final section we identify what we see as some of the more pressing issues and the more promising avenues for future work.

So, what have we learned? And what have we tried to do? The following are points we would emphasize:

• It is important to distinguish the *mechanisms* of innateness from the *content* of innateness. There need not be a one-to-one mapping between the two. We suggest that for higher-level cognitive behaviors, most domain-specific outcomes are probably achieved by domain-independent means.

•The relationship between mechanisms and behaviors is frequently nonlinear. Dramatic effects can be produced by small changes.

•What appear to be single events or behaviors may have a multiplicity of underlying causes, some of which may be distant in time.

•Knowledge ultimately refers to a specific pattern of synaptic connections in the brain. In this very precise sense, we argue that there is probably no higher-level knowledge which is innate.

•The developmental process itself lies at the heart of knowledge acquisition. Some complex behaviors may not be acquirable without passing through a developmental pathway.

•Connectionism provides a useful conceptual framework for understanding emergent form and the interaction of constraints at multiple levels; connectionism should definitely not be thought of as radical empiricism. Connectionism is still in its infancy, however, and development provides a rich set of phenomena which challenge existing technology.

•Development is a process of emergence. Connectionism provides a conceptual vehicle for studying the conditions under which emergent form arises and the ways in which emergence can be constrained.

Let us consider each of these points in turn.

A crucial point: Mechanism and content are not the same thing

In Chapter 1, we discussed innateness in terms of mechanism and content. If there is one single point we would insist on over others, it is that this distinction is crucial to understanding innateness. In our opinion, enormous confusion has been generated by failure to draw this distinction, with the result that the domain specificity of cause is conflated with the domain specificity of outcome.

In talking about **mechanisms**, we have suggested that constraints may operate at three levels: representations, architectures, and timing.

The strongest and most specific form of constraint—and the one which is most direct with regard to outcome (see discussion of "knowledge," below)—is representational, expressed at the neural level in terms of direct constraints on fine-grained patterns of cortical connectivity. This is, we think, the only kind of neural mechanism capable of implementing the claim that detailed knowledge of (for example) grammar, physics or theory of mind are innately specified.

Architectural constraints come next. In our analysis, architectural constraints can be further divided into three sublevels: unit-level architectural constraints (corresponding to the physical structure and computing properties of individual elements within a region, including neuronal types, their relative density, kinds of neurotransmitters, inhibitory vs. facilitative capacity); local architectural constraints (corresponding to the patterns of connectivity that define types of regions, and the general characteristics of layering and connectivity within a region); and global architectural constraints (corresponding to the ways in which local brain regions are connected to one another, and to the input/output pathways connecting the brain to the rest of the body).

Finally, we come to constraints on the timing of developmental events within and across regions. Time is a powerful mechanism for developmental change, and dramatic changes in species or individuals can result from small changes in the timing of developmental sequences.

As we pointed out in some detail in Chapter 5, representational constraints (the strongest form of nativism) are certainly plausible on theoretical grounds, but the last two decades of research on vertebrate brain development force us to conclude that innate specification of synaptic connectivity at the cortical level is highly unlikely. We therefore argue that representational nativism is rarely, if ever, a tenable position. This may be the single most controversial position we take in this book. To explain why mice do not become men, and vice-versa, we are probably going to have to work out a scenario involving constraints on architecture and timing. Neural networks can, we think, play a very useful role in exploring these hypotheses.

This taxonomy of mechanisms is logically independent of **content**. In particular, as we pointed out in Chapter 1, claims about innate mechanisms must be separated from the much-debated issue of domain specificity. Once we have separated the two, however, they can be used together as we try to understand how change comes about. Do we have innate mechanisms (representations, architectures, timing) that evolved in the service of specific content, including species-specific domains like language, music, faces or theory of mind? It is certainly possible, indeed likely, that uniquely human activities have played some role in the evolution of a uniquely human brain. This does not mean, however, that we are entitled to leap directly from "special" content to "special" mechanisms. When we say that a given ability is "special," we are talking about unique or unusual properties that could hold at many different levels: (a) in the structure of the task or problem to be solved, (b) in the unique solutions that we develop to deal with that task, (c) in the representations that underlie our ability to solve the problem (and hence our knowledge of the solution), (d) in the learning device and/or processing mechanism that made it possible for us to acquire those representations, and perhaps (e) in the genetic code that leads to the development of these mechanisms. The mere demonstration that a given behavior is "special," i.e., unlike anything else that we do, does not constitute evidence for (or against) the specificity of the mechanism that learns and produces this behavior. We believe that connectionist models are particularly useful as we

try to figure out just how much specificity of content we have to build into our machines, and at what level, in order to simulate something that looks like human learning.

Dramatic effects can be produced by small changes

When we see a dramatic change in behavior, we are tempted to make several inferences. First, we tend to suppose that the dramatic effect has an equally dramatic cause. We might believe that a new mechanism has kicked in (e.g., the learning of a rule), or perhaps that an existing mechanism has atrophied (e.g., the loss of a language acquisition device). Sometimes we also conclude that the cause is proximal in time, i.e., behavior changes because the child is now ready for that change.

In fact, we have offered many examples in this book of cases in which a single mechanism gives rise to very different overt behaviors at different points in time. The best known example is the well-studied case of the English past tense, where a single learning mechanism operating on a constant or incrementally expanding data base goes through dramatic changes at the behavioral level—the kind of change that was once ascribed (without further ado) to a switch from rote memory to the application of a rule. A single mechanism, operating incrementally, need not undergo abrupt *internal* changes in order to produce abrupt *external* changes in behavior. Furthermore, the sources of change may be quite distant from the changes themselves. This includes temporal distance (in cases where the causal factors long precede the changes that we ultimately observe) and logical distance (defined in terms of levels of organization). In short, things do not always happen right at the point where we notice a change. A basic lesson here is that we need to distinguish carefully between overt external behavior, and the presumed internal changes which affects behavior. The mapping may not be straightforward.

Multiplicity underlying unity: A single event has many causes, and the same event can come about in many ways

We have just argued for cases in which incremental changes in a single mechanism can lead to qualitatively different outcomes over time. This goes against an implicit assumption in much developmental research, i.e., that qualitative change at the behavioral level must reflect a qualitative shift in underlying mechanisms. One also finds a tendency in some developmental work in just the opposite direction, where unitary outcomes are attributed invariably to a unitary cause.

(Notice that if one does this, then it is very natural to infer that domain-specific outcomes have domain-specific causes. Thus, for example, if there is selective impairment in the production of regular morphology, it is tempting to suppose that the "regular morphology box" has been broken. But as we saw in Marchman's work, diffuse damage to an entire network can lead to greater deficits in processing regular forms, compared with irregulars.)

The reality is that what might be thought of as single events or behaviors can often be produced by multiple interacting mechanisms. The chick-imprinting model, described in Chapter 6, is a good example of this. The outcome in the network—imprinting on an object—requires the interaction between a nonspecific "subcortical" biasing system and a "cortical" learning system.

We have also seen that the very same event can be brought about in a variety of different ways. Noise and lesioning, for instance, can both lead to the same or similar degraded performance in networks and in humans.

On knowledge

As soon as one asks "*Where* does knowledge come from?" one has also to confront the difficult question, "*What* is knowledge, anyway?" What do we really mean when we say that a child knows something? One of the benefits of thinking in mechanistic terms

(such as the connectionist framework we have described) is that it not only requires us to be explicit about such questions, but also provides conceptual tools for answering them.

As the term is currently used in the developmental literature, "knowledge" potentially conflates different kinds of mechanisms (representations, architectures, constraints on timing) and different kinds of content (from task to gene). When a developmentalist argues that knowledge is innate (e.g., Spelke, 1994), we think it would be useful to specify exactly what this means.

We have proposed that the term "knowledge" should be used to refer to the representations that support behavior, and we have proposed operational definitions of the term "representation" that could be implemented in real brains (i.e., the fine-grained patterns of cortical activity, which in turn depend on specific patterns of synaptic connectivity) and in neural networks (i.e., patterns of activations over units in an n-layered system, which depend on the patterns of weighted connectionist between units). Under this definition, a child might still have strong constraints operating at other levels (e.g., architectural or timing) which might highly favor the emergence of specific knowledge. The knowledge itself, however, would not be innate and would require appropriate interactions to develop.

We recognize that some might reject this definition in favor of a different view, and we certainly respect that decision. But we would like to invite our colleagues to engage in a similar exercise, laying out their definition of "knowledge" in explicit, systematic terms that could be implemented in some kind of biologically plausible machine.

Why development?

We asked in Chapter 1 why development occurs at all, especially given its (apparently) maladaptive consequences (immaturity, vulnerability, dependence, consumption of parental and social resources, etc.). In fact, there are species (e.g., ungulates) which are born in a relatively mature state—they are "up and running" (liter-

ally) from the start. So it is all the more curious that so-called "higher species" should tend to have protracted periods of immaturity. And of all primates, humans take the longest to mature.

According to the perspective we have developed here, there are a number of good reasons why development should occur.

First, a long period of development allows greater time for the environment (both sociocultural and physical) to play a role in structuring the developing organism. Second, the view we have proposed is that development is the key to the problem of how to get complex behaviors (in the mature animal) from a minimal specification (in the genes). It's Nature's solution to the AI "scaling problem."

What we are not

Connectionism is sometimes viewed (usually by its critics, but sometimes by its proponents) as a return to the *tabula rasa*, i.e., to a behaviorist conception of the infant as a blank page waiting for Nurture's pen. That is, we insist, simply the wrong view. We are neither behaviorists nor radical empiricists. We have tried to point out throughout this volume not only that the *tabula rasa* approach is doomed to failure, but that in reality, all connectionist models have prior constraints of one sort or another. What we reject is *representational* nativism. Connectionism can provide an invaluable albeit incomplete set of tools for reasoning about the nature of interaction, as we set out to unravel the complex interplay of multiple internal and external constraints that underlie the diverging developmental pathways of mice and men.

Emergent form

We are also not the only nor the first to acknowledge the importance of emergent form, or to argue against a strict dichotomy between innate and learned. The new dynamic framework for studying development that we have outlined in this book represents a new contribution to an old tradition in developmental psychology and

developmental biology that includes Baldwin, Bateson, D'Arcy Thompson, Oyama, Piaget, Vygotsky, Waddington, Wimsatt, and many, many others.

Our proposals also have a great deal in common with recent writings by other developmentalist colleagues who underscore the value of dynamical systems theory as a formal implementation of the old notion of emergent form (e.g., Smith & Thelen, 1993; Thelen & Smith, 1994; van Geert, 1994; see Port & van Gelder, 1995, for a recent collection of papers on dynamical systems and cognition). Indeed, we are all engaged in a similar enterprise. What we have tried to do in this volume is to bolster these arguments with additional details from developmental neurobiology, and to provide a more concrete and precise set of proposals for the implementation of emergent form in neural networks. We also recognize that our emphasis on representations is not shared by all those who work with dynamical systems.

Models: brain or behavior

We have tried to develop an intermediate position with regard to the role of modeling, an approach which stands somewhere in between pure brain modeling and purely cognitive modeling. To be sure, for many of the higher cognitive processes of interest to us (and to most developmental psychologists), we are forced to simplify away from the details of wet brains. Our hope is that such models will embody abstract but realistic principles of learning and change that make contact with cognitive issues, while preserving some degree of neural plausibility. This is, we acknowledge, a hard choice, and we hope that future research in this area will take place at many levels from brain to cognition, with efforts to translate across levels whenever possible.

Does anyone disagree?

Does anyone really disagree with this interactionist view? Are we railing against a Straw Nativism embraced by no one but pilloried by all?

We think not. Some radical nativist proposals have been offered in the last few years, growing in strength and number in concert with rising public interest in genes for complex outcomes from cancer to divorce. To reassure worried readers that our concerns are justified, we present a sampling of statements from developmental psychologists and linguists that embody, we think, a strong version of what we have called "representational nativism." Many of these quotes come from the field of linguistics, where the Nature-Nurture debate has generated great heat and limited light. But examples from other areas are readily available.

Starting with nativist theories of cognition and conceptual change, we reiterate what we pointed out above, that the word "knowledge" is frequently used but loosely defined. If we interpret their use of the word "knowledge" in the representational sense outlined above, then the following quote from Carey and Spelke (1994) can be viewed as a manifesto in favor of representational nativism:

> *We argue that human reasoning is guided by a collection of innate domain-specific systems of knowledge. Each system is characterized by a set of core principles that define the entities covered by the domain and support reasoning about those entities. Learning, on this view, consists of an enrichment of the core principles, plus their entrenchment, along with the entrenchment of the ontology they determine. In these domains, then, we would expect cross-cultural universality: cognitive universals akin to language universals.* (p. 169)

This approach is taken several steps further in the following quote from Spelke (1994):

> *Although debates continue, studies of cognition in infancy suggest that knowledge begins to emerge early in life and constitutes part of humans' innate endowment.…* (p. 431) *Unlike later-developing knowledge, initial knowledge appears to capture what is most true about the*

> *entities that a child perceives, not what is most obvious about those entities.…* (p. 438) *If the same initial principles underlie perception and reasoning, however, then the principles could not be learned, because the child would have no other way to parse the stream of experience into the relevant entities. Initial knowledge may emerge through maturation or be triggered by experience, but learning and processing do not appear to shape it.* (p. 439)

Leslie (1992) has provided an application of this approach to the infant's "theory of mind" (i.e., beliefs about the contents of other people's minds relative to one's own—see also Leslie, 1994a,b), suggesting (a) that there is an innate knowledge system to deal with mental states, and (b) that the absence of such a system is responsible for childhood autism:

> *I have argued that the normal and rapid development of theory-of-mind knowledge depends on a specialized mechanism that allows the brain to attend to invisible mental states. Very early biological damage may prevent the normal expression of this theory-of-mind module in the developing brain, resulting in the core symptoms of autism* (p. 2)
>
> *This mechanism is essentially innate and in some sense, a specific part of the brain.* (p 20)

Leslie (1994b) places this theory-of-mind module within a more general framework that assumes innate and modular knowledge in many different domains:

> *To the extent that there are mechanisms of domain-specific development, then a deeper notion of domain is possible—one that is less software dependent, less profligate, and more revealing of the design of human cognition. This kind of domain specificity reflects the specialization of mechanisms in core cognitive architecture…the core contains heterogeneous, task-specialized subsystems. Vision is an obvious example of a specialized subsystem with a specialized internal structure. The language faculty is another* (p. 120)*…a mechanics module…the infant's processing of the physical world appears to organize rapidly around a core structure representing the arrangement of cohesive, solid, three-dimensional objects embedded in a system of mechanical relations, such as pushing, blocking, and support.* (p. 125)

Leslie's claims about innate domain-specific mechanisms seem to coincide with what we have called representational nativism.

Indeed, it would be difficult to implement such detailed claims about innate domain-specific knowledge without assuming fine-grained patterns of cortical connectivity. However, these particular authors are never explicit about the kind of neural mechanism that they envision as implementations of their nativist claims. Indeed, there may be room here for a reinterpretation in terms more compatible with the less direct but more neurologically plausible architectural and chronotopic constraints that we have put forward in this book.

In contrast with the potentially ambiguous claims offered by nativists working on cognitive development, claims within linguistics and child language are often quite explicit regarding the degree and kind of representational detail that must be innate.

In an article with the provocative title "Language acquisition in the absence of experience," Stephen Crain (1991) offers the following remarks:

> *In linguistics, one finds a near consensus on the need for constraints in explaining the development of linguistic knowledge. In addition, linguists generally find it reasonable to suppose that constraints are innate, domain-specific properties. A distinguishing feature of recent linguistic theory, at least in the tradition of generative/transformational grammar, is that it postulates universal (hence, putatively innate) principles of grammar formation, rather than characterizing the acquisition of language as the product of general cognitive growth (Chomsky, 1971; 1975). This theoretical framework is often referred to as the theory of Universal Grammar, a theory of the internal organization of the mind/brain of the language learner.*
>
> *We began by observing that recent developments in linguistic theory (the postulation of universal constraints of language acquisition), together with the observation that children's linguistic experience is quite limited (the absence of carefully sequenced input or negative evidence), reinforce the view that syntactic knowledge is in large part innately specified. What is innately given is knowledge of certain restrictions on the meanings that can be mapped onto sentences as well as restrictions on the sentences that can be used to express meanings. This knowledge is encoded in constraints. The problem for the learner is that there are no data available in the environment corresponding to the kinds of negative facts that constraints account for.*

Crain then goes into detail on the following examples, arguing that children must have innate constraints to help them infer that "He" cannot refer to the Ninja Turtle in sentence 1b.

1 a) The Ninja Turtle danced while he ate pizza.
b) He danced while the Ninja Turtle ate pizza.
c) While he danced the Ninja Turtle ate pizza.

It seems quite clear to us that innate constraints on the interpretation of examples of this kind must be very detailed indeed, requiring genetic specification of cortical wiring at the synaptic level (corresponding roughly to whatever it is we know after we have learned a language).

Lightfoot (1989) is even more explicit about the role of genetics in specifying innate linguistic knowledge, revealed in the following passages:

> *In the last thirty years, generative grammarians have been developing a selective theory of language acquisition. We have sought to ascertain what information must be available to children independently of any experience with language, in order for the eventual mature linguistic capacities to emerge on exposure to some typical "triggering experience." Cutting some corners, we have assumed that this unlearned information is genetically encoded in some fashion and we have adopted (1) as our explanatory model:*
>
> *(1) a. trigger (genotype Ÿ phenotype)*
> *b. primary linguistic data (Universal Grammar Ÿ grammar).*
>
> *The goal is to specify relevant aspects of a child's genotype so that a particular mature state will emerge when a child is exposed to a certain triggering experience, depending on whether the child is raised in, say, a Japanese or Navajo linguistic environment. (1.b) reflects the usual terminology, where "Universal Grammar" (UG) contains those aspects of the genotype directly relevant for language growth, and a "grammar" is taken to be that part of a person's mental make-up which characterizes mature linguistic capacity.…Under current formulations of linguistic theory (e.g., Chomsky, 1981), the linguistic genotype, UG, consists of principles and parameters that are "set" by some linguistic environment, just as certain receptors are "set" on exposure to a horizontal line.*

Based on the facts about brain development reviewed in Chapter 5, we think it quite unlikely that Universal Grammar could be encoded so directly in the genotype. However, many lines of evidence have been cited to support such a claim. Because language is the domain in which nativist arguments have been worked out in greatest detail, it is probably worth our while here to follow some of those arguments through, examining them from the perspective that we have outlined in this book.

Twelve arguments about innate representations, with special reference to language

Throughout the volume we have stressed that our position is not anti-nativist, but that it is essential in developmental cognitive science to (1) specify the level of innateness that is invoked for a given function (i.e., representations, architectures, timing), (2) distinguish between innateness and domain specificity, and (3) distinguish between innateness and localization. Yet one is repeatedly struck by scientists' focus on a single level (what we identify as representational nativism) in an attempt to identify genes that will explain particular behaviors. This has been particularly evident with respect to arguments in favor of a gene or a specific set of genes for language.

Let us examine the different lines of evidence that are frequently offered in favor of genetic control over the cortical representations that constitute linguistic knowledge, that is, arguments in favor of domain-specific innate representations for language. These arguments include: Species specificity, genetically based language disorders, studies of lesioned brains, activation studies of grammar in the normal brain, structural eccentricity of language, poverty of the stimulus, linguistic universals, modularity of processing, dissociations, critical periods of language learning, and robustness under different learning conditions. We invite the reader to reevaluate these claims in the light of our arguments throughout the book.

1. Species specificity

Clearly no one denies that human beings are the only creatures that are able to learn and use grammar in its full-blown form. While studies of symbol use and language-like behavior in chimpanzees and other nonhuman primates are interesting, all of the animals studied to date stop far short of the language abilities displayed by a normal human child at 3 years of age. However, it does not automatically follow that grammar itself is under strict genetic control because, as we have shown repeatedly in this volume, behavioral outcomes can be the product of multiple levels of interaction in which there is no representational prespecification. To argue for innateness, a specific link between genetic variation and some grammatical outcome must be demonstrated. Can such a link be shown to obtain in the case of language disorders?

2. Genetically based language disorders

In recent years, a number of popular reports in the press, radio and television have suggested that a link between genetic variation and grammatical outcome has indeed been found. These reports were based on preliminary work regarding a specific language impairment in one large family in London (i.e., the KE family), first documented by Hurst et al. (1990). As Hurst et al. noted, this impairment runs through the KE family in patterns that are typical of Mendelian genetic transmission (i.e., autosomal dominance). The following quote from a review by a distinguished British psychologist illustrates the attraction for many scientists of the notion that humans are special and that the human genome contains a "gene for grammar."

> *[There is] one peculiar family whose members, as a result of a defective gene, cannot form plurals, even though in all other respects they speak normally.* (Sutherland, March 7, 1993)

The startling form of highly domain-specific grammatical impairment implied by this quote would indeed (if it were true) constitute strong evidence for a genetic effect restricted to grammar.

Furthermore, given the Mendelian nature of the pattern of heritability described for the KE family, it would not be unreasonable to conclude that a single gene is responsible. The implications that such a finding would have for the nature and evolution of grammar were recognized by two eminent evolutionary biologists in a recent review entitled "The major evolutionary transitions."

> *Perhaps the most convincing evidence both for the belief that grammatical competence is to some degree independent of general learning ability, and for the possibility of functional intermediates between no grammar and perfect grammar, comes from studies of hereditary variation in linguistic competence. One remarkable case involves a family in which a so-called feature-blind dysphasia seems to be inherited in a mendelian fashion, a single dominant gene being responsible (Gopnik, Nature 344, 715, 1990). Members cannot automatically generate plurals and past tense. Although they understand the meaning of plural and past perfectly well, they have to learn each new case anew: "Paint" and "painted," "book" and "books" must be learned separately (in the case of exceptions such as "go" and "went," normal individuals must do the same). To be sure, this is not a genetical violation of one of Chomsky's rules, but it demonstrates that there can be something useful between perfect grammar and protolanguage: it also holds out the hope that we will in the future be able to dissect language genetically, as we are today dissecting development.* (Szathmary & Smith, March 16 1995; p. 231)

How sound is the evidence on which Sutherland and Szathmary / Maynard Smith base their powerful conclusions? To evaluate the actual data in some detail, it is necessary to trace the "gene for grammar" back to its original sources, and follow the relevant literature up to the present day. In doing so, several things become immediately apparent.

First, this international cause célèbre is actually based on a short and very premature report—not wholly false, but incomplete. The story begins with a brief letter to *Nature* in 1990 (Gopnik, 1990), in which a series of preliminary tests of the KE family's language abilities were reported. Gopnik concluded at the time that members of the family suffered from a deficit that is restricted primarily to the grammatical rules that underlie regular inflectional morphology (e.g., regular past-tense markings like kiss Ÿ kissed, and regular pluralizations like dog Ÿ dogs), sparing the ability to memorize words and their meanings, and the ability to memorize irregular mor-

phemes (e.g., irregular past-tense markings like go Ÿ went, and irregular pluralizations like mouse Ÿ mice). An extended version of this study was published the following year in the journal *Cognition* (Gopnik & Crago, 1991). Based on their data for the KE family and other dysphasic subjects in Montreal, Gopnik and Crago conclude with the following:

> *It is not unreasonable to entertain an interim hypothesis that a single dominant gene controls for those mechanisms that result in a child's ability to construct the paradigms that constitute morphology.* (p. 47)

Although these conclusions were couched in terms of tentative hypotheses and have since been weakened as more comprehensive data on both the language and nonlinguistic cognition of this family became available (Gopnik et al., in press), the attractions of the Gopnik & Crago data were rapidly taken up by others in the field. For example, in a review article in *Science*, Pinker (1991) argued for a neurally and perhaps genetically based double dissociation between regular and irregular grammatical morphemes. Pinker contrasts the dissociation in the KE family (where regulars are supposedly more impaired than irregulars) with a complementary deficit reported in Williams Syndrome. Integrating this apparent double dissociation with other arguments for a distinction between regular and irregular forms (see Chapter 3 for more details), Pinker marshalled the following argument:

> *Focusing on a single rule of grammar, we find evidence for a system that is modular, independent of real-world meaning, non associative (unaffected by frequency and similarity), sensitive to abstract formal distinctions (for example, root versus derived, noun versus verb), more sophisticated than the kinds of rules that are explicitly taught, developing on a schedule not timed by environmental input, organized by principles that could not have been learned, possibly with a distinct neural substrate and genetic basis.* (p. 253)

Of course, if the data were as clear-cut as originally supposed, then Pinker's claims would indeed constitute a strong case for the domain-specific innateness of the human language faculty (see, also, Pinker, 1994b). Although Pinker makes it clear that he does not believe in a single gene for grammar, he does argue that the specific

profile reported for the KE family reflects a specific genetic program for language that cannot be derived by simply reconfiguring less specific cognitive systems. Despite Pinker's own disclaimer of a single gene for grammar, reviews of his influential book in popular newspapers and magazines brought the grammar gene concept back to center stage in its original form.

A sharply contrasting view of the KE family comes from a group of British researchers at The Great Ormond Street Hospital for Children in London, where the family has been under study since 1988. No one disputes the conclusion that this disorder has a genetic base, a finding that was first reported by Hurst et al. (1990), who conclude that the pattern of inheritance in the KE family may be due to a single gene (i.e., autosomal dominance). The contrast between Gopnik and the Great Ormond Street data revolves not around the innateness of this disorder, but its domain specificity. In a letter of rebuttal to Gopnik and her colleagues published in *Nature,* Vargha-Khadem and Passingham (1990) note that the grammatical symptoms displayed by members of the KE family are just one part of a much broader symptom complex:

> *Gopnik has focused on only one aspect of the disorder, and it is inaccurate to conclude that the speech and language problem stems simply from an impairment in manipulating feature markers.* (Vargha-Khadem & Passingham, p. 226; see also Fletcher, 1990)

In 1995, a comprehensive report of the London group's results was published, including extensive tests of language and nonlanguage abilities (Vargha-Khadem et al., 1995). Controlling for age, they showed that affected members of the KE family perform significantly worse than unaffected members on all but one of 13 language tests (including tests of phonology, grammar and lexical abilities). In addition, the affected members are significantly lower on *both* verbal and nonverbal IQ (about 20 points in both cases). Moreover, they perform very poorly on a nonlanguage test of oral-facial praxis (i.e., production and imitation of simple and complex movements of the tongue and mouth). In fact, their oral/facial apraxia is so severe that some members of the family are completely unintelligible to strangers. In contrast to the claims that have been

made in some of the scientific literature and in the popular press, the affected members of the KE family are significantly worse than the unaffected members on *both* regular and irregular morphology, with no evidence for a dissociation between the two. The disparity between the Vargha-Khadem et al. and the Gopnik and Crago findings is explained by the fact that the former are based on a much larger set of experimental morphological contrasts than the original four past-tense items used by Gopnik and her colleagues: Two regulars (kissed and walked) and two irregulars (went and was). Yet these subsequent clarifications about the broad array of deficits actually present have gone largely unnoticed in public discussions of the KE family.

The search for a genetic base to grammar has been prominent among some developmental investigators too (e.g., Matthews, 1994; Rice, in press; van der Lely, 1994). Consider the syndrome known as Specific Language Impairment, or SLI. SLI is usually defined in terms of expressive (and perhaps receptive) language abilities being one or more standard deviations below the mean, and below the same child's nonverbal Performance IQ, in the absence of evidence for mental retardation, frank neurological impairment, abnormal hearing, social-emotional disorders and/or social-demographic conditions that could explain the disparity between language and other cognitive functions (Bishop, 1992; Bishop & Rosenbloom, 1987). Hence, by definition, SLI is thought to represent an impairment that affects language and nothing else. Furthermore, it has been repeatedly demonstrated that grammatical morphology is particularly vulnerable in children with SLI (Johnston & Kamhi, 1984; Leonard, 1992), albeit with little evidence for the claim that regular morphemes are more vulnerable than irregulars (see Marchman, Wulfeck, & Weismer, 1995, for a discussion of this point). Most important for our purposes here, a number of studies have shown that SLI and associated disorders (especially dyslexia) tend to run in families (Bishop, 1992; Bishop & Rosenbloom, 1987; Pennington, 1991; Tallal, Ross, & Curtiss, 1989; Tallal et al., 1991).

At face value, this looks like evidence for genetic specification of grammar. However, as with the case of the KE family, detailed

studies of children with SLI have shown that the syndrome is not restricted to language, nor to grammar within language.

For instance, Tallal and her colleagues have amassed a large and compelling body of evidence suggesting that children with SLI suffer from a deficit in the processing of rapid temporal sequences of auditory and (perhaps) visual stimuli (Tallal, 1988; Tallal, Stark, & Mellitts, 1985). Other specific nonlinguistic deficits implicated in SLI include symbolic play in younger children (Thal et al., 1991), aspects of spatial imagery in older children (Johnston, 1994), specific aspects of nonverbal attention (Townsend et al., 1995), and a wide range of neurological soft signs (e.g., Trauner et al., 1995). More recently Tallal and colleagues (Tallal et al., 1995) have shown that by short-term training of SLI children on oral language in which phonemic transitions are lengthened to facilitate processing, dramatic improvements are found at all levels of the language system, suggesting once again that it is not morphosyntax that is specifically impaired, but some more general deficit in the processing of rapid sequential material.

In fact, problems with receptive processing of grammatical morphemes are observed in an even wider range of populations, including anomic aphasics who display no grammatical deficits in their spoken language (Bates, Devescovi et al., 1994; Devescovi et al., in press), and a subset of elderly patients who are hospitalized for nonneurological disorders (Bates, Friederici, & Wulfeck, 1987). Receptive agrammatism has even been induced in college students who are forced to process sentences under stress (e.g., perceptual degradation and/or cognitive overload: Bates, Devescovi et al., 1994; Kilborn, 1991; Miyake, Carpenter, & Just, 1994).

It appears, then, that grammatical morphology is selectively vulnerable under a wide range of conditions, genetic and environmental. A parsimonious account of all these findings would be that grammatical morphology is a weak link in the processing chain of auditory input, one that is highly likely to be impaired when things go awry. None of these examples points necessarily to specific genes for grammar.

Why did the 1990 Gopnik letter to *Nature* find such a wide audience while the 1990 rebuttals by Vargha-Khadem and Passingham

and Fletcher, and their subsequent full-scale report, fall on deaf ears? Why is the notion of specific "genes for grammar" so attractive? There seems to be a deep-rooted desire to believe that humans are not only unique (every species is, after all) but that our uniqueness arises from a quantum leap in evolution. The grammar gene(s) satisfies this desire. In reality, the fact that we share almost 100% of our genes with nonlinguistic species suggests that our language capacity is more likely to be the not-so-simple result of multiple constraints operating at different levels in ways argued throughout this book.

Localization

As stressed in Chapter 5, localization and innateness are not the same thing. For example, studies using positron emission tomography (PET) have revealed specific areas of the brain that are very active in response to real words and to nonsense words that follow the rules of English spelling, but not to words that violate English spelling rules (Petersen et al., 1988). Although this clearly suggests localization, surely all would agree that a prespecified brain circuit for English spelling is a highly unlikely candidate for genetic determination. In the same vein, PET studies have revealed specific areas of the brains of chess masters that glow at specific points in the game (i.e., Nichelli et al., 1994), and yet no one (least of all Nichelli et al.) would argue that chess is innate, at least not in any nontrivial sense. These localized areas for spelling or chess are the *result* of progressive specialization following massive experience, not of innate specifications. Yet evidence of localization is often used to buttress claims about innateness, on the assumption that innate systems have inherited their own dedicated neural architecture (Fodor, 1983; Pinker, 1994b). Kandel, Schwartz, and Jessell (1995) express these assumptions in a chapter on brain and language:

> *Chomsky postulated that the brain must have an organ of language, unique to humans, that can combine a finite set of words into an infinite number of sentences. This capability, he argued, must be innate and not learned, since children speak and understand novel combinations*

> *of words they have not previously heard. Children must therefore have built into their brain a universal grammar, a plan shared by the grammars of all natural languages.* (p. 639)

Does the human brain contain at birth specific neural regions that are dedicated exclusively to the representation and processing of grammar? Two complementary lines of evidence are relevant to this question: studies of language ability in adults and children with focal brain injury, and studies of language-related brain activity in normal individuals.

3. Localization I: Lesion studies

It has been known for some time that aphasic disorders in adults are strongly correlated with lesions to the left hemisphere in right-handed individuals. This is not in dispute. But during the 1970's and early 1980's, a stronger claim was made: Deficits in grammar are associated with lesions to specific areas of left frontal cortex, resulting in agrammatic Broca's aphasia (Caramazza & Berndt, 1985; Heilman & Scholes, 1976; Zurif & Caramazza, 1976; for an extension of this assumption to early child language, see Greenfield, 1991).

A key to this argument lies in the generality of the agrammatic deficit, in both comprehension and production. This hypothesized syndrome would include the well-documented impairments of expressive language that characterize Broca's aphasia (i.e., telegraphic speech with omission of inflections and function words), as well as the receptive processing of the same grammatical elements—a receptive deficit that can only be detected when the patient is prevented from using semantic information to interpret sentences.

For example, Broca's aphasics typically perform quite well in the interpretation of passive sentences like "The apple was eaten by the boy," but they perform poorly on semantically reversible sentences like "The boy was chased by the girl," sentences that force these patients to rely exclusively on grammatical information. Based on findings like this, it was proposed that agrammatic Broca's aphasics have lost all access to grammatical knowledge,

with the sparing of semantics. A complementary analysis was offered to explain the comprehension deficits, empty speech and severe word-finding deficits associated with fluent Wernicke's aphasia: All these symptoms could be interpreted in terms of a by-product of a central deficit in semantics, with the sparing of grammar.

The proposed double dissociation between grammar and semantics seemed to unite a broad array of symptoms under a single description, while providing neurological evidence for a modular distinction between grammar and semantics that had been proposed on independent grounds within generative linguistics. This appealing theory has, however, been challenged by subsequent data, including (1) case studies showing that some agrammatic Broca's aphasics perform normally on receptive grammar tasks (Miceli et al., 1983; Tyler, 1992), (2) a large number of studies showing that receptive grammatical deficit is in fact observed in many different clinical groups and, as mentioned above, in normal adults who are forced to process sentences under adverse conditions (Bates, Devescovi et al., 1994; Blackwell & Bates, 1995; Kilborn, 1991; Miyake, Carpenter, & Just, 1994); (3) the finding that "classic" agrammatic patients are able to make fine-grained judgments of grammaticality (Linebarger, Schwartz & Saffran, 1983; Shankweiler et al., 1989; Wulfeck, 1988); and (4) cross-linguistic studies showing that so-called agrammatic aphasics retain detailed features of their native grammar, evident in many different aspects of their expressive and receptive language (Bates (Ed.), 1991; Menn & Obler, 1990).

At this point in the history of aphasia research, few investigators still cling to the idea that grammatical knowledge is localized in or around Broca's area. Instead, the grammatical deficits of nonfluent patients are usually explained with reference to processing deficits that are only indirectly related to grammar itself. This transition in thinking about the nature of agrammatism is illustrated in the following quote from Edgar Zurif, one of the original proponents of central agrammatism in the 1970's (Zurif et al., 1993):

> *The brain region implicated in Broca's aphasia is **not** the locus of syntactic representations per se. Rather, we suggest that this region provides processing resources that sustain one or more of the fixed*

> *operating characteristics of the lexical processing system characteristics that are, in turn, necessary for building syntactic representations in real time.* (p. 462)

The story is rendered still more complex by a growing body of evidence showing that children with left-hemisphere injury are able to acquire language abilities within the normal range (Aram, 1992; Eisele & Aram, 1995; Reilly, Bates, & Marchman, in press; Vargha-Khadem et al., 1991, 1992). As we noted in Chapter 5, this does not mean that the brain is equipotential for language at birth, because specific correlations between language deficits and lesion site are observed in the first stages of language development (Bates, Thal, et al., 1994; Reilly et al., in press; Thal et al., 1991). However, these correlations do not map onto the lesion-symptom relations that are typically observed in adults (e.g., comprehension deficits in children with right-hemisphere injury; expressive deficits that are specifically associated with the putative receptive areas of left temporal cortex). Furthermore, there is no evidence for a dissociation between grammar and semantics in these children. The same left temporal lesions that lead to a delay in expressive vocabulary are also implicated in the delays that are observed during the emergence of grammar.

Although it is possible that a specific grammar region may emerge from lesion studies at some future time, the evidence to date suggests that grammatical knowledge is broadly distributed in the adult brain. Indeed, recent evidence from split-brain patients suggests that detailed judgments of grammaticality can be made in the right hemisphere (Baynes, 1990; Baynes & Gazzaniga, 1987; Zaidel, 1990), suggesting that grammatical knowledge is distributed across both hemispheres.

4. Localization II: Activation studies of grammar in the normal brain

Tools for the study of language-associated brain activity include positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and event-related scalp potentials (ERP). All of these methods have been applied to the study of processing at

the single-word level (Garnsey, 1993; Petersen et al., 1988), and PET and ERP have also been used in studies that attempt to disentangle semantic and grammatical processing (Hagoort et al., 1993; King & Kutas, 1992; Mazoyer et al., 1993; Mecklinger et al., 1995; Münte, Heinze, & Mangun, 1993; Neville et al., 1991; Osterhout & Holcomb, 1993). Although most studies show greater activation in the left hemisphere during language tasks, there is little or no consensus across studies regarding the regions within the left hemisphere that have greatest responsibility for phonological, semantic and/or grammatical processing (Poeppel, in press).

An initial round of ERP studies comparing the brain potentials associated with semantic vs. grammatical violations did lead some investigators to propose that grammatical violations are associated with specific patterns of brain activity (Hagoort et al., 1993; Mecklinger et al., 1995; Neville et al., 1991; Osterhout & Holcomb, 1993). Candidate syntactic potentials have included an early negative wave that is largest over left anterior scalp (i.e., the so-called N280—Neville et al., 1991), and a slow positive wave that extends to around 500-800 msec following a grammatical error (i.e., the syntactic positive shift, SPS or P600—Hagoort et al., 1993; Osterhout & Holcomb, 1993). However, more recent studies have shown that the N280 is neither specific to function words nor specific to grammatical/syntactic errors. Instead, this negative component is present for all words, with a latency that reflects the word's length and frequency of usage (King & Kutas, 1995). In fact, both the N280 and the N400 appear to lie along a continuum that cuts across word class.

In the same vein, studies varying the probability of violations (i.e., the percentage of items that contain such a violation) have shown that the P600 behaves in many respects like a well-known positive component (the P300) that is observed with violations in many different verbal and nonverbal tasks (Coulson, King, & Kutas, 1995). Certain types of grammatical (morphosyntactic) violations elicit both an N400 and a P600 simultaneously (Kluender & Kutas, 1993, in press; Münte et al., 1993). In addition, one other component (the left anterior negativity or LAN, presumably different from the N280 although it overlaps spatially and temporally) that has been linked to syntactic violations has been observed even in the absence

of violations, specifically in the context of perfectly grammatical sentences (wh- questions) that make heavy demands on working memory (Kluender & Kutas, 1993, in press). Both the multiplicity of "grammar" and "semantic waves" and the substantial overlap between them indicate that grammar and semantics are not each uniquely associated with "signature" components of the ERP. Thus, while it is fair to say that different violations are associated with different patterns of brain activity, these differences do not constitute evidence for a special, autonomous grammatical processor.

In fact, the number of so-called language-specific areas are multiplying almost on a daily basis. Every new functional imaging study seems to bring another language area to our attention. This is true not only for PET, fMRI, and ERP, but also for studies of brain activity in epileptic patients with electrode grids placed directly on the cortical surface. For example, new language areas have been discovered in the fusiform gyrus, in left and right basal temporal cortex (Lüders et al., 1986, 1991), an area that has never previously been linked to language impairments in lesion studies. But it is important to stress that this proliferation of areas is not unique to language; it has been occurring in every area of perception and cognition as fineness of technical measures improves (e.g., Orban et al., 1995).

This all leads to the conclusion that domains like language do not live within well-defined borders, at birth or at any other point in development. Existing evidence for the localization of language (such as it is) provides little support for the idea that children are born with neural mechanisms that are prespecified for and dedicated solely to language processing.

5. Structural eccentricity

Beyond the biological/neurological arguments, investigators of a strong Nativist persuasion have also invoked the structural eccentricity of language to argue for its domain-specific innateness. Language, it is argued, is so different from other behavioral systems that it cannot be explained by general mechanisms of learning and cognition. It is true that language involves a host of opaque and

arbitrary structural facts, a point that Chomsky has made in numerous publications. There is, for example, no obvious communicative account of the fact that:

*We expected Bill to like each other.

is ungrammatical whereas

Each of us expected Bill to like the other(s).

is grammatical, since both sentences are perfectly comprehensible (Chomsky, 1975, p. 101). Why is the fact that language is often expressed in structures which have no obvious communicative advantage relevant to the innateness debate? It is relevant because eccentricity of form is used as evidence for eccentricity in the underlying processors responsible for the learning and processing of those forms. And yet we have offered numerous examples throughout this volume in which peculiar input-output mappings were learned by a connectionist network with no prespecified representations. Furthermore, such eccentricities may arise for reasons which have nothing to do specifically with domain-specific (linguistic) representational constraints, but rather with peculiarities at the level of architecture or timing which may be domain neutral in their content (but still having specific consequences for different domains). We conclude that the mere existence of structural eccentricities tells us little about either the level of the constraint which gives rise to them, or their domain-specific or domain-general character.

6. Poverty of the stimulus

Perhaps the most compelling behavioral evidence for innateness comes when a newborn gazelle leaps to its feet and runs at birth, or when a spider makes a perfect web on the first try with no prior opportunity to observe web-weaving by another spider. Such examples constitute evidence for performance in the absence of experience and learning. Human language does not fall directly within that class of behaviors, because children take at least three years to

get a full-blown grammar up and running. However, a cornerstone of strong nativist claims about language rests on the idea that grammars cannot be learned, no matter how long it takes, because children do not receive enough evidence from the input to support the kinds of generalizations that they ultimately draw. This is, of course, the well-known poverty-of-the-stimulus argument.

Gold's theorem (Gold, 1967) has been used as the foundation for arguments of this kind, a formal proof that grammars of a certain class are unlearnable in the absence of negative evidence (i.e., in the absence of evidence that certain sentences are illegal in the language). But of course, as we have noted elsewhere (e.g., Chapters 2 and 3), Gold's theorem rests on some very strong assumptions about the nature of grammar and the nature of the learning device, e.g., (1) that the grammar to be learned consists of strings of discrete symbols governed by discrete rules, (2) that the learning device itself is an hypothesis-testing device that makes yes/no decisions on each trial, and (3) that whole grammars are tested one at a time against each incoming sentence (for detailed discussions, see Pinker, 1979; Wexler & Culicover, 1980). If these assumptions do not hold (and they are implausible for human children), then Gold's theorem is simply not germane. Connectionist simulations of language learning can be viewed as empirical tests of learnability claims, based on very different assumptions about the nature of grammatical knowledge and the nature of the learning device. Some of these simulations have already shown that the impossible is in principle possible, although a lot of important work obviously still lies before us.

7. Universals

Another common argument for innateness revolves around the existence of language universals. Despite their differences, the grammars of natural languages have many properties in common. Because these universals could not have arisen by chance, they are offered as prima facie evidence for innate grammatical knowledge (Roeper, 1988).

However, in this volume we have seen a number of examples of connectionist networks that discover similar and stable solutions to well-defined problems; an example we have used often is XOR. Another such example comes from models of the visual system. It is very common for units that have center-surround and line orientation receptive fields to emerge in simulations in which 3-D images are mapped onto a 2-D array, and the task for the network is to reconstruct the original 3-D object from the lower dimensional representation (e.g., Lehky & Sejnowski, 1988). These solutions do not "look like" the problem they solve, but they are ubiquitous, if not universal. The are solutions which are contained in the structure of the problem space—as long as the processing mechanism has the appropriate constraints for solving the problem.

A similar story may underlie the universals that have been reported for grammar. The grammars of natural languages may be thought of as solutions to an even more daunting dimension reduction problem, in which multi-dimensional meanings must be mapped onto a linear (one-dimensional) output channel (the mouth). The fact that these grammars may not always obviously resemble or reflect the underlying content of the message may be irrelevant to the question of where these solutions come from.

8. Modularity of processing

In the mature speaker/listener, certain aspects of language processing take place rapidly and efficiently, and they often appear to be impervious to contextual factors and/or conscious strategies. These processing characteristics are the hallmarks of modularity as it is defined by Fodor (1983), and the existence of modularity has been used by many theorists as another form of evidence for the innateness of "special purpose" processors (e.g., Gardner, 1983).

However, as Fodor himself has noted (e.g., Fodor, 1985), the same processing characteristics have been demonstrated in laboratory studies of perceptual-motor learning. Any arbitrary skill can achieve "automaticity" if it is practised often enough, under highly predictable conditions (Posner & Snyder, 1975; Shiffrin & Schneider, 1977) —which means that the skill becomes very fast, efficient, hard

to think about, and impervious to interference once it gets underway. Hence modularity may be the result of learning rather than its cause, i.e., "Modules are made, not born" (Bates, Bretherton, & Snyder, 1988, p. 284), a process of progressive modularization, not initial modularity (Karmiloff-Smith, 1986, 1992a).

9. Dissociations

We have already pointed out that localization and innateness are logically and empirically independent. Therefore, arguments based on the dissociations that result from focal brain injury are not necessarily germane to the innateness debate. However, it is sometimes argued that dissociations give us insights into the organization of "virtual architecture" or "mental organization," independent of the neural substrate (Marshall, 1984; Shallice, 1988). In the same vein, developmental dissociations have been used to argue for the "natural" or "innate" boundaries of the mind (Cromer, 1974; Gardner, 1983; Smith & Tsimpli, 1995). To the extent that language can be dissociated from other cognitive systems, it can be argued that language is a "natural kind."

Although this is a plausible argument (see Karmiloff-Smith, 1992a, Chapter 2, for an extended discussion), dissociations do not necessarily reflect innate or domain-specific boundaries, because they could result from deficits that are indirectly related to the final outcome. This is clear from a number of connectionist simulations in which a general-purpose learning device "acquires" some aspect of language and then suffers a "lesion" to the network (Hinton & Shallice, 1991; Marchman, 1993; Martin et al., 1994; Plaut, 1995). These simulations often display dissociations that are strikingly similar to those reported in the literature, and this in the absence of prespecified (innate) structure and / or localization of domain-specific content. In other words, dissociations can be explained in a number of ways. Although they provide potentially useful information about the seams and joints of the resulting cognition, they do not alone constitute compelling evidence for innateness.

10. Maturational course

Fodor (1983) has argued that biologically relevant and domain-specific modules like language follow their own peculiar maturational course. It is true that language has its own characteristic onset time, sequencing, and error types, and the course of language learning includes many examples of U-shaped functions and discontinuous change. Does this constitute evidence for genetically timed maturation? Only if one assumes (as some theorists do) that learning is linear, continuous and subject to immense variability in onset time and sequencing—a claim that is increasingly difficult to defend. We have furnished many examples throughout this volume of nonlinear learning in neural networks, including discontinuous shifts, U-shaped functions, novel overgeneralizations and recovery from overgeneralization. These phenomena have been shown to occur in the absence of prespecified representations, although their emergence always depends on the nature of the initial architecture, the nature of the input, and the time course of learning (cf. Chapters 3, 4 and 6). The line between learning and maturation is not obvious in any behavioral domain (see especially Chapter 5). Language is no exception in this regard.

11. Critical periods

This is a special case of the maturational argument above, referring to constraints on the period in development in which a particular kind of learning can and must take place. The quintessential examples come from the phenomenon called "imprinting," which is argued to differ from domain-general learning in four ways: stimulus specificity, absence of external reinforcers, irreversibility, and temporal boundedness. Although the original imprinting phenomena have been challenged, imprinting persists as a metaphor for constraints on language learning. Three kinds of evidence have been invoked to support the idea that there is a critical period for language learning in humans: "wild child" cases (e.g., deprived children like Genie, or Izard's Wild Boy (Curtiss, 1977; Lane, 1976),

age-related limits on recovery from brain injury (Lenneberg, 1967), and age-related limits on second-language learning (Johnson & Newport, 1989).

Although we agree that there are age-related changes in the ability to learn and process a first language, in our opinion these changes are not specific solely to language (e.g., they occur for many other complex skills introduced at different points in life). Furthermore, we have shown that changes of this kind may be the result of learning rather than its cause. Within a nonlinear neural network, learning results in concrete changes in the structure of the network itself. Across the course of learning (with or without the appropriate inputs for a given task), the weights within a network become committed to a particular configuration (a process that may include the elimination of many individual elements). After this "point of no return," the network can no longer revert to its original state, and plasticity for new tasks is lost. Marchman (1993) has demonstrated this kind of critical-period effect for networks lesioned at various points in grammatical learning, and the same explanation may be available for other critical-period phenomena as well. What is important is that the seeming critical-period does not require (although it does not preclude) an extrinsic, genetically driven change in learning potential. It has also recently been shown that the so-called critical period even for second-language learning may turn out to involve differences in critical types and amounts of experience rather than actual age of acquisition (Bialystok & Hakuta, 1995).

12. Robustness

Normal children rush headlong into the task of language learning, with a passion that suggests (to some investigators) an innate preparation for that task. Several lines of evidence attest to the robustness of language under all but the most extreme forms of environmental deprivation (Curtiss, 1977; Sachs et al., 1981). Examples include the emergence of "home sign" (rudimentary linguistic systems that emerge in the gestures produced by deaf children of hearing parents—Goldin-Meadow & Mylander, 1984) and creoliza-

tion (the transformation of a pidgin code into a fully grammaticized language through intermarriage and constant use of the pidgin in natural situations, including first-language acquisition by children—Bickerton, 1981; Sankoff, 1980). A particularly striking example of creolization comes from a recent study of the emergence of a new sign language among deaf adults and children who were brought together in a single community in Nicaragua less than a dozen years ago (Kegl, Senghas, & Coppola, 1995). It has been argued that the rapid and robust emergence of grammar with limited input constitutes solid evidence for the operation of an innate and well-structured bioprogram for grammar (Bickerton, 1981). We would agree that these phenomena are extremely interesting, and that they attest to a robust drive among human beings to communicate their thoughts as rapidly and efficiently as possible. However, these phenomena do not require a preformationist scenario (i.e., a situation in which the grammar emerges because it was innately specified). We argued in Chapter 1 and above that grammars may constitute the class of possible solutions to the problem of mapping nonlinear thoughts onto a highly constrained linear channel. If children develop a robust drive to solve this problem, and are born with processing tools to solve it, then the rest may simply follow because it is the natural solution to that particular mapping process.

* * *

We end this section with a brief comment on the issue of the use of the term "innate" and social responsibility. At this writing, interest in innate ideas and innate constraints on cognition has reached another high-water mark. This is evident in popular books on "human instincts" (e.g., Pinker, 1994b), but it is also evident in books that argue for racial differences in intelligence (Herrnstein & Murray, 1994).

Of course, these two approaches to innateness are not the same. One can obviously argue for the innate basis of characteristics shared by all human beings while rejecting the notion that individual or subgroup differences are immutable. But this neat division runs into difficulty as we move from behavioral description to the

elucidation of an underlying mechanism. The problem is that genetic differences and genetic commonalities come from the same source. If we ascribe a complex and highly specific ability to some direct genetic base, then we have opened the door for genetic variation and the disturbing sociopolitical implications that ensue.

As scientists, we cannot and should not hide from facts. If our data force us to an unpopular or unhappy conclusion, then we must live with the consequences, playing out our moral beliefs on the plane where such beliefs belong (i.e., in civic responsibility and political action). At the moment, cognitive scientists and neuroscientists have a long way to go before our findings achieve the status of immutable and irrefutable truth.

However, the words we use to explain our interim findings still have important consequences. The word "innate" means different things to different people, some refusing to use it at all. To some people, "innateness" means that the outcome in question cannot and should not be changed.

We disagree with such a position and have taken pains throughout this volume to be as explicit as possible about the multiple levels of complex interactions at which something might be innate. If scientists use words like "instinct" and "innateness" in reference to human abilities, then we have a moral responsibility to be very clear and explicit about what we mean, to avoid our conclusions being interpreted in rigid nativist ways by political institutions. Throughout we have stressed that we are not anti-nativist, but that we deem it essential to specify at precisely what level we are talking when we use terms like "innate." If our careless, underspecified choice of words inadvertently does damage to future generations of children, we cannot turn with innocent outrage to the judge and say "But your Honor, I didn't realize the word was loaded."

Where do we go from here?

Connectionism is not a theory of development as such. Rather, it is a tool for modeling and testing developmental hypotheses. We hope to have convinced readers during the course of this volume that connectionism provides a rich, precise, powerful and biologically plausible framework for exploring development, and in particular for rethinking the complex issue of innateness. We are excited about the possibilities for new ways of understanding development which are opened up by the framework we have outlined here. But we are also painfully aware of how little we understand and how much remains to be studied. It is with neither cynicism nor discouragement that we acknowledge what is usually true in science: At any given point in time it will be the case that we only know 5% of what we want to know, and 95% of that will eventually turn out to be wrong. What matters is that we might be on the right track, that we are willing to discard those ideas which are proven wrong, and that we continue searching for better ideas. In this final section we would like to identify what we see as particularly important challenges. Thus, this section represents for us a program for the future.

Multi-tasking in complex environments

Some of the most influential early connectionist models of psychological phenomena were motivated by a desire to understand how interactions between various knowledge sources and how context might affect processing (e.g., the word-reading model of McClelland & Rumelhart, 1981). And connectionism is often seen as being quite opposed to modularist theories (although as we pointed out in Chapter 2, there is no necessary reason why connectionist models should not be modular; the more interesting question is what the content and the ontogeny of the modules is).

It is therefore ironic that most current models are in fact highly task-specific and single-purpose (see discussion in Karmiloff-Smith, 1992a, Chapter 8). Most of the past-tense verb-learning models, for instance, do not attempt to integrate the knowledge of past-tense morphology with other parts of the morphological system. Yet

clearly the child does not learn language according to some scheme whereby Monday is past tense, Tuesday is relative clauses, etc. Not only is language learned in an integrated fashion, but language learning is highly dependent on the acquisition of behaviors in non-verbal domains.

We believe that in order to study the development of complex behaviors, it will be crucial to have models which have greater developmental and ecological plausibility. These must be models which are capable of carrying out multiple behaviors (in computer jargon, multi-tasking) in complex environments which require behaviors which are coordinated and integrated.

Active and goal-oriented models

Most models are passive. They exist in an environment over which they have no control and are spoon-fed a preprogrammed diet of experiences.

Babies, on the other hand are active. (They may be spoon-fed, but consider how often the spoon ends up in their hair or on the floor.) To a large extent, they select their environment by choosing what they will attend to and even where they will be. Thus, there is an important distinction between what is input (presented to a child) and what is uptake (processed). Sometimes this difference is not under control (as when a child's physical or cognitive level of maturity precludes processing an input), but at other times it is.

Furthermore, children have agendas. Their behaviors are typically goal-oriented; they do something (or learn something) because it moves them closer to some goal. Thus if we want to understand why some behaviors develop sooner or later, it is often important to understand the goals towards which the behaviors are directed. Contrast this with the typical network, which has no internal drives or focus.

Where do these goals come from? Who writes the child's agenda? It seems to us that these are not solely developmental questions and that to answer them we must also attend to phylogenetic (i.e., evolutionary) considerations. We are encouraged by the recent interest (not only among connectionists) in artificial life and evolu-

tionary models. We do not believe that the many basic goal-oriented behaviors are learned or taught. Rather, they are evolved. So these new evolutionary models may eventually be profitable avenues for studying ontogenetic development.

Social models

From an early age, the infant's behaviors are highly social. Many of the earliest behaviors are related to the infant's desire to interact with its caretakers. Later in life, social interactions play a critical role in shaping and stimulating the child's development.

In a similar vein, we would like to see models which have a more realistic social ecology. There are many behaviors which have strong social components. Language is an obvious case: It is conventional (i.e., relies on socially agreed-upon forms), and it is one of the most useful currencies for social interaction. There are also phenomena such as the development of a child's awareness that others have private mental states which cannot be directly observed (Theory of Mind), which cannot be modeled by networks in isolation.

Recent work in cognitive science may provide a good theoretical basis for developing social connectionist networks. Hutchins' "distributed cognition" asserts that many cognitive phenomena occur at the level of group interactions. Thus, in just the same way that one cannot understand what makes a termite nest function without looking at behaviors in the aggregate (individual termites are fairly dumb when considered in isolation), there are many human behaviors which are manifest at the level of group activity. Hutchins himself has developed models of communities of networks (Hutchins & Hazlehurst, 1991), and we see this as a fruitful direction for future developmental work.

Higher-level cognition

The models we have discussed throughout the book do not touch on planning, reasoning, and theory building. Indeed, Karmiloff-Smith has argued that connectionist models account particularly well for some aspects of development, but stop short of addressing

important developmental facts regarding higher cognitive functions (Karmiloff-Smith, 1992a, 1992c). It is well documented that children go beyond successful behavioral outcomes to form theories about how, for instance, language and the physical world function. Future models will have to focus on ways in which connectionist simulations that embody implicit knowledge about tasks could eventually create explicit knowledge that could be transported from one domain to another in the service of reasoning and theory building. We do not believe there is any reason in principle not to believe that such models are possible—but their present lack presents an important challenge for the future.

More realistic brain models

We stated at the outset that we take biological plausibility as an important source of constraint on our models. This does not require that we limit ourselves to what is currently known about the biology, simply because too much remains unknown. And models can play a useful role in motivating the search for new empirical data. But we also take seriously the charge that models not stray too far from what is at least plausible (if not known), and our models must stay abreast of new developments.

The search for greater realism must necessarily be carried out with a greater understanding into what *matters.* Brains are clearly very complicated things. But it is often not clear what part of that complexity contributes directly to function, and what part is simply there as an outcome of evolution's necessarily Rube Goldberg approach—the result of random variations which have proven to be adaptive, but which do not necessarily make for an optimal or even tidy system. One challenge is the need to identify the relevant features of brains, and then find a level of modeling which continues to make contact with the anatomy and physiology, but which can also make contact with behavior.

A second challenge is the development of models of the brain which take into account the *entire* (well, at least several parts!) of the brain. The O'Reilly and Johnson model of imprinting which we dis-

cussed in Chapter 6 is a good example of the type of model we have in mind.

A final note

This brings us to our end, but obviously we also regard this as a beginning. For the six of us, the journey which we have followed to get to this point has been exciting, stimulating, and often personally challenging.

This is not a book which any one of us could have written alone. And we are certainly different people as a consequence of having written it together. We have not always agreed at the outset of our discussions, but we see this as a strength. Each of us comes with a different background and frequently a different perspective. These differences, we feel, have played an important role in making this book what it is. Rather than trying to submerge disagreements, we have respected our differences of opinion, and have attempted to forge a synthesis which goes beyond the initial differences to a new perspective which reconciles them.

We hope that you, the reader, may feel some part of the excitement which we feel at the new prospects for understanding just what it is that makes us human, and how we get to be that way.