Evolution and intelligence: beyond the argument from design.

Terrence W. Deacon,
Biological Anthropology
Boston University, Boston, MA 02215

From A. Scheibel & Wm. Schopf, eds. (1997) **"The Origin and Evolution of Intelligence,"** *Jones and Bartlett publishers, pp. 103-135. Revised post-publication April 2000 (originally written in 1995-6).*

The persistence of top-down explanations in biology

When the theory of natural selection was first presented to the scholars of the last century, many found it to be too implausible to believe. The incredulity of many great thinkers at the time, from brilliant biologists to articulate theologians, was based on a well-reasoned common sense understanding of the world: Left to chance, things tend to get less organized, not more. Millennia prior to Darwin, this same reasoning led Aristotle to criticize the natural philosophy of his contemporary, Empedocles, who argued that all natural processes are the actions of blind chance and that organisms arise out of the preservation of useful accidents (see Aristotle's Physics). Aristotle easily found innumerable examples of end-directed design in nature that he felt could on no account be explained from such a minimalist perspective. But Aristotle was wrong about this, and only after more than twenty centuries of musing about this conundrum, did scientists come to realize the power of the opposed conception for explaining biological phenomena. When the logic behind Empedocles' insight was rediscovered and given a more substantive interpretation by Darwin and Wallace, it revolutionized biology by providing an answer to this counterintuitive problem. This has become widely appreciated, not just by biologists, but by the general lay public educated in basic biology.

Nevertheless, despite the wide acceptance of natural selection as the major source of evolved adaptations, it remains a very counterintuitive notion. In a famous early 19th century monograph on Natural Theology, William Paley (1802) argued that it was beyond credulity to suggest that something as complex as a pocket-watch, with its precisely aligned and interdependent gears and movement could have been organized by anything other than an intelligent designer. Not just its complexity, but the apparent purposive (i.e. functional) design of its parts would mark it as a product of intelligence and forethought. Since organisms appear far more complex and intricate in design and function than a pocket watch, Paley argued that it is even more absurd to imagine that they could have arisen without the guiding hand of an intelligent designer. Darwinism, however, offered a theoretical apparatus to explain how, despite great improbability, it might nonetheless be possible to explain the spontaneous production of complex adaptation in entirely non-purposive terms. Though even a century later there are still nonscientists who blindly doubt the facts of biological evolution, the efficacy of the selection process that Darwin proposed to account for it cannot be doubted. It has become a standard laboratory tool in molecular biology, information processing, and even engineering, used to spontaneously generate adaptive solutions to complex problems. Even if we were to discover that it was not, as we suspect, the major formative principle at work in shaping the course of life on earth, this would not change the fact that it is the only serious contender for a natural process capable of generating such elaborate and finely adapted structures. So unquestionable is this mechanism, that in a modern inversion of Paley's claim, many biologists now argue that

whenever we encounter adaptive complexity in nature the default hypothesis must be that it is a product of natural selection.

And yet Paley's perspective has not been totally eradicated from biological explanations. A core feature of this view is in fact central to much biological thinking, despite the fact that no serious biologist can now doubt the role of Darwinian processes in the evolution of organisms. It is the thorough rooting out of this idea, which I believe to be a major source of confusion about the nature of biological processes even for professional biologists, that is the purpose of this essay.

Paley's original insight was motivated by the analogy of a living organism to a precisely designed machine: a watch. The Darwinian response has been to deny the necessity of intervention by intelligent planning and construction in order to produce such complex organization. In the modern view, Paley's designer has been replaced by natural selection. But the implicit comparison of an organism to a machine has remained. The neoDarwinian view is that organisms are put together *as if* they were designed. The products of evolution are still conceived of as clockworks where each part has been fashioned with respect to its contribution to the function of the whole organism. Their adaptations are, in a sense, conceived as being accidentally designed with respect to some purpose, such as foraging, predator evasion, or reproduction.

This engineering analogy comes naturally. Designers often find themselves looking to nature for new solutions to structural problems, and everything from molecular diffusion systems and artificial neural networks have borrowed "design principles" from living counterparts. It is difficult to avoid the habit of thinking of organism structures and functions in engineering terms, given the elegance of some of nature's solutions. The aerodynamics of bird wings, the streamlining of fish and dolphin bodies, the ball and socket structure of shoulder and hip joints, or the countercurrent exchange arrangement of blood and water flow in fish gills, each have an engineering elegance that human inventions have struggled to emulate or have serendipitously converged upon. So one might defend the engineering analogy as at least a heuristic convenience, especially since engineering analyses can also yield insights about organism functions. Unfortunately, such analyses can also be misleading, and this becomes more of a problem as the complexity of the system increases as in the case of brains and intelligent behaviors.

The view of natural selection as a designer and organisms as its designed products is an eliminative program. Design is replaced by the retention of useful accidents. Purposive function is replaced by mechanistic operations that are merely correlated with a result, not chosen to achieve it. Intentional adjustment to the demands posed by the environment is replaced by the blind implementation of prespecified behavioral and developmental programs encoded in genes. Hence, Richard Dawkins (1976), in his paradigm-challenging book *The Selfish Gene*, asks us to imagine multicellular organism bodies as replication machines—great lumbering robots designed by genes to ferry these genes from one generation to the next. And, though cautioning his readers to take his anthropomorphic turn of phrase, when speaking of "selfish genes," as merely a rhetorical convenience, he succeeds in conveying this counterintuitive perspective in part by cleverly inverting the typical view of organisms as purposive and genes as mere strings of molecules. He beguiles his readers into abandoning the natural teleological perspective at one level by replacing it with a provisional, apparently merely metaphoric use of it at another. Thus, ultimately, this line of reasoning suggests that the one presumed exception to this *post hoc* logic, our own rational intelligent purposive behavior, may not be so rational or purposive after all.

Intelligence might simply be understood as the evolution of robotic programming of a particularly complex and high level of flexibility, but a pre-designed program nonetheless. The presumed conclusion appears to be that such behaviors and experiences are ultimately reducible to collections of preprogrammed mechanisms in which all but superficial recombinations of pre-specified elements are prefigured. This offers a final eliminative step. Even the apparent intentionality and purposiveness of thought processes appear to be reducible to a form of predetermined mechanism.

This conceptual short cut, embodied in the engineering model of organisms, is a source of misleading intuitions about how living systems are assembled and how they operate—a short cut that continues to lead otherwise lucid biological analyses astray. This is because it allows us to ignore how the differences in the means for achieving functional adaptation have contributed to the forms they produce. Beginning from this perspective, we are implicitly invited to imagine that the plans for organism designs are passed from generation to generation in coded sequences along the strands of DNA that make up each species' genomes and that these plans are decoded and used to determine the realization of these designs during ontogeny, like construction workers following a blueprint. Organism design is imagined to be embodied in a set of genetic assembly instructions that have been written and debugged by natural selection (as opposed to an all intelligent designer).

Yet we know that evolution produces organisms according to a very different *modus operandi* than does the purposive design of useful tools and other devices. Not only do these two very different processes arrive at solutions to the same problems in very different ways, they inevitably produce very different kinds of solutions as well. One of the main reasons for this is that the assembly of organisms is autonomous or self-determined, rather than under extrinsic control, and this imposes very different design constraints than when the assembly process and assembly instructions are independently conveyed and executed by a designer. Both methods for producing adaptation may yield superficial resemblances in response to common problems and yet hide radically different logics of design and function. The question is, "What is this difference in design logic?"

What I hope to demonstrate in this essay is that the design logic we tend to use to describe living adaptations often exactly inverts the reality, and its assumption can be as misleading as Paley's view was for understanding the nature of organism evolution. This does not mean that I wish to dispute the Darwinian interpretation of the origins of adaptation. Just the contrary. I want to propose a more far-reaching and radical Darwinism, that replaces design thinking with self-organization and selection thinking at all levels, including organism function. Surprisingly, I think this ultimately leads to a non-eliminative theory of design, function, and intelligence in biology.

The ontogenetic information bottleneck

A common misconception often implicit in evolutionary arguments is that all the adaptive variety that an organism exhibits in its body functions and behaviors can be explained as a consequence of information imparted through genetic inheritance (with the exception of a little fine tuning added by learning). Changes in gene frequencies produced by natural selection are indeed capable of generating adaptations of remarkable power and subtlety, but genetic change is by no means the only way that adaptive biological structures and processes can be produced. Recent critics of strong selectionist theories, such as Stephen J. Gould, in an effort to respond to the incompleteness of the gene selection models of evolutionary processes, have suggested that

we consider that some of this adaptation occurs at higher levels in evolution, such as at the population, species, and higher taxonomic levels (see for example Gould and Lewontin's 1979 critique of adaptationism). He also points out that chance historical contingencies may play a rather major role as well. Most evolutionary theorists, however, suggest that the conditions under which species level selection could contribute a significant adaptive result are quite restrictive and generally unrealistic for most populations. Other alternative explanations for the origins of order in organism design have considered the role that intrinsic systemic constraints play in determining systemic regularities. For example, Stuart Kauffman (1992) has shown how complex patterning can emerge spontaneously simply due to the finite relationship within networks of interacting elements, like genes. These constraints on the range of possible patterns, introduce biases into evolution that may have had a significant effect on the ways genes determine cell types and body structures. The extent to which these can account for a large fraction of the patterns of organism design and evolution is still unknown. But I want to argue that these do not exhaust the possibilities, and that all evolutionary adaptation accounts that are restricted to an analysis of gene frequencies and structural/functional consequences are missing an essential *middle level* of explanation.

One major limitation on the power of population genetic selection processes to produce adaptive variety in body structures is that the genome is a fairly inefficient mechanism for transmitting "intelligent" adaptations. By intelligent adaptations I mean those that involve highly flexible, indirect, complicated or unobvious solutions to adaptive problems. Genetic transmission is a limited transmission medium because cellular molecular transmission of information is constrained by distances, rates, and specificity of the chemical processes that underlie it. There are as a result tradeoffs in both information density and in packet size. These tradeoffs translate into upper limits on genome and cell size, constraints on metabolism and reproductive rates, and on the size of the fraction of genes which carry structural information. There are reasons to suspect that the genomes of vertebrates are near the upper limits. These limitations need to be considered in the context of the information required for adaptation. With each level of structure (molecular-cellular-structural-functional) in a complex multicellular body there is an exponential increase in the amount of information necessary to control its construction and operation. In addition, the range of possible adaptive responses also grows exponentially with the complexity of the organism, including the complexity of its genome. As a result, a solution-by-solution, mutation-by-mutation, trial-and-error search for adaptive responses conducted at the lowest level of design information (genes) is simply incapable of adequately sampling a useful range of alternatives (see also, Kauffman, 1992). Finally, with increasing design complexity, it becomes almost certain that random changes in structure will result in decline of function, not enhancement. This produces increasing inflexibility. Fiddling with even a single connection in a television or digital computer inevitably degrades function and usually produces catastrophic results. Theoretical evolutionary genetics has assumed that such an incalculably minimal likelihood of useful errors in a complex system can ultimately be explained away by appealing to the law of large numbers: vast periods of evolutionary time and huge numbers of individuals.

I am suspicious that in accepting a narrow view of Darwinism-restricted-to-genes we have ignored its role in numerous other levels of supportive processes, and that these are essential for a full account of how gene-selection can succeed in arriving at adaptation. Even if these astronomical statistics could bear the weight of this multidimensional adaptive "search" across the many levels of variables involved, such a means for achieving complex adaptation

would be irrelevant in the face of a more information efficient means. Natural selection will favor any mechanism that reduces this massive sampling problem and can generate the requisite information with less information transmission overhead. Instead of relying on immense time and intense selection to achieve immensely improbable modifications of genes that just happen to fit each new adaptive problem, organisms have evolved ways to circumvent this need wherever possible. Those lineages that evolved ways to effectively "off load" this information generation and transmission demand to a number of subsidiary systems, inevitably supplanted those for which time and numbers were the only recourse.

For generations, biologists conceived of development and evolution as additive processes. New adaptations, new structures, and new functions added to previous ones, producing more and more complicated organisms from the bacterium to the bat, and from the zygote to the full-grown zebra. But this view of evolution and development is an anachronism. With very few exceptions, the history of life is characterized by experiments with diverse rearrangements of the same few basic body plans. On average, vertebrates have larger genomes than species in most other animal phyla (though most of it is non-coding DNA that few single celled organisms can afford to allow to accumulate), but among vertebrates there is no trend suggesting that species with larger more complex bodies and brains require more genes to produce them. Most estimates place the number of genes coding for protein products in mammals like mice and humans at a number that is less than 100,000 and this is probably a reasonable estimate for vertebrates in general since mammals have genome sizes in the middle range for vertebrates. This is particularly significant for brain evolution in vertebrates. First, it suggests that few if any additional genes may be involved in the production of huge human brains as opposed to tiny dwarf salamander brains. Second, considering that human brains probably contain over a hundred billion neurons with thousands of connections each, it suggests that the wiring instructions that could possibly be specified genetically are probably insufficient by many orders of magnitude to wire brains, connection by connection. And of course, the extent of under-determination of connectivity grows geometrically with size. This means that human brains, which are among the largest on the planet, must be among the least pre-specified. These are all reasons to suspect that the genetic instructions for building brains are not likely squandered on the micromanagement of individual neural connections, and probably not even on the specification of tens of thousands of connections.

This brings up an even more enigmatic problem of brain design. If brain circuits are not well determined by genetic instructions, how is it possible that precise and predictable capabilities arise in evolution? How else could highly heritable behavioral capabilities evolve? In fact, brains are remarkably similar from individual to individual within a species, and within different species within the same orders, even when they differ greatly in size. Despite this growing information shortfall, large brains, like human brains, are organized into many more distinguishable subdivisions than small brains and are as well suited to operate their big bodies as are mouse brains for mouse bodies. Where does all the additional information come from, if not from genes?

The answer requires a Darwinian explanation, but not a phylogenetic one. What I mean by this is that the process of brain development itself resembles a Darwinian process in certain important respects. The reason for this is clear. Darwinian processes can spontaneously generate new useful information where none previously existed, and they need little pre-specification to get underway. It is only natural, then, that this immense information gap in ontogeny should get filled in by the action of a kind of within-the-body natural selection, to

sample information that is implicitly present outside the genome. In this regard, replication and growth of connections are the analogues of organism reproduction, competition for connection sites is the analogue of competition for resources and mates, and how well particular connections match the signal processing requirements of the body becomes the analogue of natural selection, in this microcosm of evolution. Neural connection patterns that are appropriate to manage a complex body evolve to assume complementary niches of information processing much as species evolve competitively and cooperatively to fill the complementary niches in an ecosystem (though the analogy is only partial, since ecosystems do not have their own consolidated means of replication). This evolution-like process for building brains allows brains to become adapted to the bodies they inhabit as well as to their own internal constraints with a minimum of preplanned design (see also Deacon, 1995).
occur later.

Darwinian physiology

Richard Dawkins takes the Darwinian alternative to its logical extreme by arguing that Darwinism is the only known mechanism for spontaneously producing adaptive complexity (Dawkins, 1986). But there are two ways to interpret this claim. The first interpretation is what I have been describing as the standard evolutionary doctrine, i.e. that all complex living processes are the results of biological evolution and shaped in their designs by Darwinian selection. But there is a second, more radical interpretation possible: that the complex adaptive process that occur within and during the building of an organism are also themselves Darwinian-like processes, and not merely its results. The difference in this second interpretation is that it implicitly suggests that the details of complex adaptive responses are very minimally prefigured in genetic instructions, and rather that they are created online, so to speak, as the demand arises and at the corresponding level of organism structure. From this perspective there is no need for a full plan for the organism to be encoded in the genome. What is required is something like the codes for a series of biasing mechanisms capable of coaxing otherwise spontaneous processes down predictable pathways. Dawkins mostly intended the first interpretation. This is borne out by his mechanistic account of organisms as mostly preprogrammed robots. But the two views are not at all at odds. The more encompassing perspective, requires the traditional view as its limiting case. It might be caricatured by the phrase "Darwinism all the way up," by which I mean that partially independent nested Darwinian processes generate adaptive complexity at all levels of organism self-assembly and function.

The general utility of the selectionist model beyond topics in the phylogeny of life, has been widely recognized. Extensive research has already shown that selection processes are the modus operandi for cellular-molecular mechanisms underlying underlying immune response, gamete production, and aspects of neural development. However, even among biologists, few regularly consider how the basic elements of this explanation of spontaneously self-ordering processes apply to less remote aspects of organism function and what the broader significance of this logic of undesigned design might be. The reason for the ubiquitous role of evolution-like processes in widespread aspects of organism function is the generic nature of the conditions that can produce Darwinian processes, minimal conditions that can arise in a variety of contexts beyond population socioecology and genetics. These minimal conditions include: a) the existence of self-replicating units of information whose replication is subject to low levels of spontaneous replication "errors" that arise independent of any function, b) competition for resources that limit or enhance replication of these units, and c) consequent selective bias in

replication rate and probability with respect to the differential functional consequences of these variations.

The power of natural selection theory to explain the origins of adaptive complexity derives from its simplicity and inevitability given certain minimal conditions. It is no more than the statistics affecting replicative processes in the context of entropy. As the end of the 20th century approaches the scientific community has not only come to recognize the phylogenetic importance of this statistical mechanics of life, but also is beginning to appreciate its ubiquity at every level of living function. Its widespread importance is a consequence of one simple fact: Darwinian selection is capable of spontaneously generating novel adaptive structure, and of achieving this by generating new information where none previously existed (unlike matter-energy, information can be created and destroyed). This is in fact what phylogenetic evolution demonstrates. Organisms take advantage of this fact by hierarchically embedding one Darwinian processes within another in order to build their adaptations. And it provides a powerful way for organisms to amplify adaptive capacity and extend it into more complex and unpredictable realms. Such hierarchic embedding occurs spontaneously as higher order competitive-replicative processes are coöpted by lower order ones. Within an organism, competition between gametes, selective replication of immune cells, and competition between growing axons for growth factors and synapses in the developing brain all utilize variants of Darwinian processes to approach optimal adaptation and bridge the gap between gene-level processes and organism-level processes of selection. And there may be many more levels to this than we now recognize. For example, many geneticists are beginning to suspect that the majority of gene-gene interactions that control development and the molecule-molecule interactions that regulate cell metabolism also, at base, depend upon this same sort of statistical mechanics of replication and competitive selection to produce complex adaptive effects.

The widespread employment of Darwinian processes in developmental and physiological processes derives from their ability to provide an autonomous source of adaptive responses. As both simulations and organic evolution demonstrate, they can arrive at nearly optimal solutions to adaptive problems without pre-specification or monitoring of outcomes. And the less that must be specified in detail the better, or rather the more likely they are to be employed. Biological "solutions" arrived at by incorporating information that is spontaneously and reliably present will inevitably succeed over those that are dependent on information that must be passed on in an entirely pre-specified form because systems that require detailed element-by-element design specifications will be more susceptible to entropic influences, less flexible to perturbation and changing conditions, and more improbable to modify usefully in reasonable periods of time. In contrast, organisms that have made the best use of spontaneous Darwinian mechanisms will inevitably evolve faster, out-reproducing and out-competing those with more top-down design strategies. Moreover, hitting upon such a strategy is not unlikely. Darwinian processes arise spontaneously under a range of conditions that are common in living systems where replicative processes are ubiquitous. So, the underlying evolutionary logic never even needs to be specified or designed into an organism. It merely needs to be designed around, coöpted and constrained by providing subtle biases to help shape the replication and selection processes involved.

This view of organism "design" complicates how we think about comparative anatomy in a number of ways. Consider, for example, the notion of homologous organs (e.g. finger tips and hooves, or the cerebellum in mammal and frog brains). In classic theories it was possible to distinguish between equating organs due to common descent of design information (homology) and due to common function due to convergence or parallelism (analogy). But this distinction

becomes a problem when we begin to consider ontogenetic Darwinian processes. The similarities between brains and brain systems in different individuals and different species must thus be the result of a process analogous to convergent evolution, where highly similar selection forces act on homologous structures to produce similar organs. The more directly homologous the substrates and the more highly constrained the selection conditions the more similar will the end product appear. Consider vertebrate pectoral fins and wings. Sharks, lobe finned fishes, plesiosaurs, and dolphins all converged on essentially the same fin "design" from slightly different starting points (ray-finned fishes arrived at a similar solution but using a different construction material). And pterosours, bats, and flying lemurs converged on essentially the same wing "design" logic (birds arrived at a similar solution using a different construction material). The design information, so to speak, was implicit in the preconditions of selection. The more similar these are, the more similar the result. The similarity of selection preconditions in development is, however, a distributed relationship among a great many body systems. This contributes a great deal of informational redundancy. The result is that, for the most part, incremental changes in one element will have a subtle and seldom catastrophic effect on the whole. Indeed, one of the recent supports for this hypothesis comes from studies with so-called "knock-out" breeds, in which geneticists inactivate a specific gene and see what happens when an organism homozygous for this knock-out tries to develop. Often, even with apparently very important genes, knock-out organisms develop to resemble normal individuals because other systems have found ways to compensate, in part, for what was missing (real advances are however made when we discover cases where there is a major re-organizational consequence). There appears to be a sort of hierarchy to this plasticity, in which events that occur early in development (e.g. genes turned on early) whose effects will tend to ramify in all later stages, are more highly conserved (which implies that they are under stronger genetic selection) than those that occur later.

It is the purpose of this presentation to explore the logic of this evolutionary tendency for building organisms as evolutionary processes within evolutionary processes, by drawing on examples from brain evolution and brain development. This most complex product of evolution is the ultimate expression of amplification of adaptation, so the expression of this logic should be most explicit when it comes to brains. It is also where our intuitions about design and purpose and mechanism most easily become confused. It is precisely where mechanistic and purposive models come into direct conflict, and where common sense intuitions find both of these accounts to be wanting. These qualities make it an ideal place to start to trace such a thoroughly Darwinian logic of organism design. *Ultimately, the problem of explaining the apparent intelligence behind evolutionary adaptations and of explaining the evolution of intelligence as an adaptation can be seen as the same problem, and in a general sense have similar explanations.* As we are beginning to see more serious attempts to discover the links between brain evolution, brain development, cognitive processes, it becomes increasingly important to appreciate the related processes at intermediate levels between genetic evolution and cognition.

Brains adapt to bodies

If brains were designed the way we design watches or computers, flexibility from generation to generation, and evolutionary adaptation in the long run would be a near impossibility. The brains of large mammals, like ourselves, are some of the most complex objects that have ever existed. With so many millions of interconnected interdependent parts, significant modifications of one component would need to be correlated with complementary

modifications of innumerable other interdependent components in order to avoid catastrophic disruption of system-wide functions. This is one very good reason why no one recommends redesigning computers or television sets by randomly modifying connections or parts in millions of devices and testing to see which "mutations" work best. Aside from the obvious waste involved, the chances that any of the millions of possible changes would produce enhanced function are essentially zero. Technological progress is the result of redesign where the engineers involved must pay very close attention to all the detailed ways that their innovations interact with one another and the other parts of the device. Redesigning and upgrading complex systems such as microprocessors and jet airliners, requires tens of thousands of engineers contributing thousands of hours of checking and rechecking both how their own parts work and how they interact with those designed by others. And even then surprises tend to emerge when the whole system is first assembled. And these designs are no where near as complex as even simple organisms and their brains. Living things aren't like designed devices because that approach was far too cumbersome to ever have evolved.

So let's take a closer look at brain development with an eye toward identifying the levels of Darwinian processes involved. In order to predict the outcome of a Darwinian process, one needs to know two things: 1) the sources of bias affecting the growth and replication of the basic units of selection (i.e. neural connections) and 2) the invariant features of the context in which this growth occurs that provide the competitive milieu and selection biases that determine which of these units will develop and persist and which will not. The parallels are not difficult to recognize in brain development.

Starting with the process of cell production, we can see clear evidence of selection processes at many levels of the nervous system (for pioneering reviews see Purves and Lichtman, 1980, and Cowan et al., 1984). Some of the first examples discovered involved the production of neurons directly controlling muscles: motor neurons (see review in Purves, 1988). During development, it was noticed, these output neurons were produced in greater abundance that what was seen in mature individuals. Sympathetic ganglia, whose neurons projected to the smooth muscles of the viscera, and spinal motor neurons that projected outputs onto limb muscles, seemed to go through a culling process as the organism matured. In frogs and chicks, where it was easier to experiment on early embryos, it was discovered that the extent of this culling could actually be increased or decreased by modifying the peripheral organs to which they projected axons. By removing their targets more cells were induced to die off and by grafting additional organs (e.g. a supernumery limb) fewer cells were eliminated. Apparently, these cells were initially overproduced and then found themselves in competition for resources somehow provided by the peripheral target structures. These resources turn out to be both molecular (like analogues to food) and structural (like analogues to protective burrows, etc.).

Early in development the output branches of motor neurons (axons) grow somewhat exuberantly and nonspecifically and end up overlap one another on the same muscle cells. Competition ensues to make contacts (synapses) with these cells in which in the end only one axon will win, and only those with stable synapses seem to provide molecular signals necessary to keep the source neuron alive. Those branches of axons that fail to establish connections die back, so to speak, and those cells who fail to establish any stable connection will die altogether. Selection mediated cell death, then, turns out to provide a precise mechanism for matching the sizes and distribution of populations of neurons to the sizes and distribution of muscle masses in the rest of the body. From an evolutionary point of view, there need be no correlated change in neural cell production or cell distribution to match changes in muscle size and distribution that

have resulted from selection for different modes of locomotion (Deacon, 1990; Katz and Lasik, 1983; Purves, 1988; Wilczynski, 1984).

The same logic turns out to be utilized throughout the developing brain, and not just for motor systems, but for sensory systems and even intrinsic systems as well. Consider the visual system, for example. In different vertebrates the direction that the the two eyes face may differ by almost 180° (as in many fish and hoofed mammals) to almost 0° (as in owls and humans). In species where the visual fields overlap there is the possibility of using the nearly redundant information to aid in depth perception, but given the rang in possible overlap one might suspect that it needs to be accomplished differently in different brains. Indeed, the way these connections map onto the visual analyzers in the brain is quite complex. As is the case for many sensori-motor systems, the visual projections into the central analyzers of the brain maintain a topographic organization (though somewhat distorted, as are many world map projections). In animals with binocular overlap (like monkeys and ourselves) each half retina views most of the same visual field as the same side half retina of the other eye. The projections into the brain of an adult split according to their visual field of origin, so that the parts of the eye that view the left visual field cross over in the midline on their way into the brain and both map onto the cerebral cortex (the folded gray matter sheet that covers each half of the forebrain) on the right side.

What is remarkable about these maps is that they are both separated and overlapping in a complicated way in the visual cortex. It is as though each map was cut into meandering strips (like zebra stripes) and then put together into a single map, interposing each sides' strips between the other's, so that points on a strip that represent the same point in viewed space are aligned right next to one another. This allows neurons nearby one another to compare signals, and thereby extract depth information from the slight shifts in disparity that result from the way distance influences the convergence or divergence of lines of sight. One might imagine that such a complicated map organization, requiring such precise alignment, would require very detailed pre-specification. Evidence to the contrary was discovered over two decades ago when it first became possible to trace the course of individual input connections at different stages in development. What was found was an early rather messy pattern of projections in which the two eyes' maps sort of diffused into one another with poor point-to-point precision. Axons tended to branch and fan out in overlapping patterns in the visual cortex. But during development the degree of overlap and fan-out is reduced via competition between axons, and many of what in hindsight we might call misdirected and nonspecific branches are selectively eliminated to produce the final precisely sculpted pattern. Much more has been learned since then about the level of initial biases that help axons approximate their initial trajectories and about the nature and extent of competitive processes involved, but the basic logic underlying all is a Darwinian logic.

The power of selection processes to produce such a precise complex pattern has been most forcefully demonstrated by a parallel pattern produced in the frog visual system, but in this case by a very abnormal experimental manipulation. Law and Constantine-Paton (1981) grafted an additional (third) eye onto frog embryos' heads in order to study the effect on axon growth to the brain. In frogs, the major visual analyzer is not the cortex (frogs don't exactly have a brain structure that is comparable to cerebral cortex) but rather a paired midbrain structure called the optic tectum (which is also a sheet-like structure), which contains the visual maps. Like other amphibians, reptiles, and fish, frogs eyes don't view any significant degree of overlapping visual space and each retina projects its map in total to opposite sides of the optic tectum. The triclops frogs, however, do experience a lot of visual overlap. The third eye, placed above and adjacent

to one of the existing eyes, views a large part of the same region of space. When the experimenters examined the projection pattern in the optic tectum they discovered that these two eyes projected interdigitated stripe maps, analogous to the mammal binocular maps, in this very different brain structure. What is so remarkable about this is that never before in evolution have there been frogs with binocular vision of this sort; never before has there been a need for a mechanism to precisely interdigitate visual maps. It emerged spontaneously without evolutionary precedent precisely because of a similarity in the processes of selection. It is an expression of convergent developmental selection, analogous to the convergent evolution of fins and wings, only producing much more precise convergence from seemingly very different origins.

These examples beg the question of exactly how the competitive processes work. The answer is only incompletely known, and involves many mechanisms. Most agree on two critical components: 1) there are growth factors that are provided by the recipient (post-synaptic) cells which are necessary to maintain patent connections and for which axons compete, and 2) what determines whether or not an axon will be able to hold onto a synaptic connection with respect to competing axons vying for the same target cell has something to do with correlating activity patterns that favor those axons staying connected to the same target cell which tend to fire in synchrony with one another and the target cell as well (Purves, 1988). The end result is that subtle spatial and timing biases in initial connectivity contribute to the relative synchrony or dyssynchrony of signals converging on any particular target, and these biases become amplified by the progressive action of millions of signals and selective elimination events that ensue. So some initially rather subtle biases, that can be controlled developmentally by some rather generic growth processes, are capable of adjusting the development of both large scale and micro-scale connectional patterning, and in the end produce intricate appropriately detailed circuits to match.

It now appears that this general principle is at work at all levels of brain development. In fact, the remarkable pattern of maps for different sensori-motor modalities and submodalities that divides up the cerebral cortex of mammals, probably is also only loosely biased by genetic design and yet comes to exhibit a remarkable inter-individual and cross-species consistency. Studies by O'Leary and colleagues (see review in O'Leary, 1992), for example, have shown that transplanting immature cerebral cortex from one position to another in the cerebral mantle of rats does not cause the later appearing sensori-motor maps to track the repositioning. Rather, they develop input and output connections that are appropriate for where they end up (Stanfield and O'Leary, 1985). The reason for this was uncovered by these same workers in studies that showed that early cortical outputs are quite nonspecific with respect to their targets (O'Leary and Stanfield, 1989). All areas of cortex initially send outputs to nearly all types of cortical targets, mostly be way of extra branching of their axons. Later in development these branches are pruned so that visual areas only project to subcortical visual processors, auditory areas only project to auditory processors, and motor areas only project to motor structures. So it doesn't so much matter where in cortex the projections originate.

The same is true for inputs to cortex that determine what information each sector will process. Growing chunks of fetal cortex and thalamus (the major source of inputs to cortex) in tissue culture has shown that it doesn't matter what combinations of cortical and thalamic regions are grown together, they each will interconnect as well as any others (e.g. see studies by Molnár and Blakemore, 1991; and Yamamoto et al., 1992). Moreover, studies in which growth of developing inputs to thalamus from various sensory systems have been interrupted early in development, demonstrate that the remaining inputs can recruit what thalamic regions (nuclei)

were supposed to be the targets for the missing inputs (e.g. see studies by Frost and Metin, 1985; and Sur et al., 1988). And this shift in thalamic input patterns is passed on to the cortical representation of these nuclei as well. The cortical tactile auditory region can, for example, become visually responsive due to visual inputs taking over the auditory nucleus of the thalamus (the medial geniculate nucleus). There is even a curious natural experiment that demonstrates this. The blind mole rat (*Spalax*), has only vestigial eyes. In its brain the visual nucleus (the lateral geniculate nucleus) instead gets most of its input from auditory subcortical sources, and there is a corresponding shifting of functional boundaries in the rest of the thalamus and the recipient areas of cortex to match (Doron and Wollberg, 1994). Where other rodents have visual cortex, the blind mole rat has auditory and somatic sensory cortex, not because the one was eliminated and the others added in any modular sense, but because in the competition for space the displacements of connection patterns at lower levels produced ramifying effects at all others. Thus cortical areas' inputs and outputs are both competitively determined along with the patterns of connections within cortex as well. It is a pattern generation process that is entirely systemic and distributed.

Brain transplants

One consequence of the fact that most of the information for wiring brains is produced, not inherited, in each generation, is that most aspects of species brain differences are likely not specified in any detail. The very same genetic and molecular information can serve very different purposes not only in different parts of the brain but in very different brains as well. Some of the most convincing evidence for this cross-species generality comes from experiments in which chimeric brains are produced by transplantation techniques. Immature neural cells, harvested prior to the point in development where neurons have extended fragile axons and dendrites, can be transplanted from one brain to another, both across ages (from fetal to adult) and across species. By placing immature neurons from one stage or species into another it is possible to experimentally probe for the relative influence of intrinsic and extrinsic signals for development.

As part of an effort to develop alternative fetal cell sources for transplantation treatments for neurodegenerative diseases, I and my colleagues [principally Ole Isacson, see references] at McLean Hospital have studied the growth of fetal cells from pig brains in the brains of adult rats and monkeys (and have recently had the opportunity to follow this in human clinical trials as well). The Darwinian features of neural developmental processes had given us reason to predict that the signals that controlled neural maturation and connectivity in these different species would likely not be all that dissimilar, and so might allow cross-species transplants to function nearly as well as same species transplants (so long as their immune rejection could be held at bay; which was accomplished with the same sorts of drugs used for other types of organ transplants). It also gave us a chance to observe exactly how the different species' cells would interact.

One interesting discovery, was that the developmental clock, that decides when cells move on to succeeding stages of differentiation and growth, seems to be intrinsically controlled. Pig cells mature at the same rate whether in pig brains or rat brains. This turned out to be a boon, both for the experiments and for understanding features relevant to their clinical application. Adult brains are highly resistant to axonal growth. This is one reason that brain damage is seldom repaired spontaneously. We still do not know why these inhibitory mechanisms are activated, but they slow axon growth by an order of magnitude at least. This

meant that fetal rat neurons implanted into an adult rat brain often can't grow their axons to reach appropriate targets. They reach the end of their growing period without establishing synapses and lacking appropriate support, the axons and sometimes the cells die. But pig neurons stay immature and growing more than five times as long as do rat neurons, and as a result, even at their slow rate of growth through the stubbornly resistant adult brain tissue, pig axons could grow to reach targets that rat axons could not (Deacon et al., 1994; Isacson et al., 1995). Remarkably, pig neurons taken from one region of the fetal pig brain and implanted into an adult rat brain grew their axons to the rat brain target structures that were homologous to the normal pig brain targets for those neurons. Cortex cells grew to striatum and midbrain, striatal cells grew to midbrain but not cortex, and midbrain cells grew to striatum and cortex, and so on (Isacson and Deacon, 1996). Growth into these targets did not appear to be topographical, but rather generic, and could occur even if the transplants were placed in odd sites within the host brain, and most importantly, these newly established connections could replace functions in rat who had lost these functions due to loss of some of their own neurons. It is this reason that we have been able to approach human clinical therapy using pig cells with some confidence that it will work, and indeed early results are quite promising.

The implications of this for evolution are that even some incredibly species-specific neural functions, such as echolocation or language processing, have probably been achieved not by specific circuit changes encoded by specific genes for those functions but by using the same old developmental information, slightly modified by systemic effects and generic biasing of global brain relationships. Pig cells might even be playing a role in supporting improved language functions in human patients, even as you are reading these words! For this reason, understanding how such a uniquely human function arose in the first place may not require any highly specific genetic explanation. There may be no intrinsically pre-specified language circuits in the human brain, only circuits that have inherited this function as a consequence of a doubly Darwinian process.

Does this mean that there is no modularity and no pre-specificity at any level in brains? Is the whole affair a Darwinian competitive free for all? Though I believe that the determination of organism "design" is Darwinian all the way down (even to the molecular level), this does not preclude the evolution of highly localized modular systems, nor does it exclude the possibility of significant species and individual differences that have a precise genetic basis. However, understanding that these too arise from Darwinian mechanisms during development may help us to gain a clearer understanding of the conditions that must be met for such specificity to result.

Size plays a critical role. Many small worms and related creatures (e.g. leeches) have nervous systems in which the neuron by neuron identity and wiring are highly predictable, so much so that it is possible to individually name (or number) specific neurons in specific ganglia and brain regions. At small body sizes, individual neurons are much more like separate organs subserving distinct functions and this precludes the kind of statistical specification that goes on in species where each neural "organ" is composed of anywhere from hundreds to billions of neurons all processing similar information. But even in very small species with countable neurons, it seems to be the case that programmed cell death and systemic selective processes are at work, they are just more highly and easily constrained by their context.

The level at which higher level brain functions are specified in different species has recently been studied directly, by putting whole chunks of developing brains into other species to replace corresponding chunks removed from them. One of the most interesting examples of this experimental approach has used developing chicks and quails. It turns out to be possible to

remove a section of the neural tube in a very young Japanese quail embryo and replace it for the correspondingly removed section of a chick neural tube, so long as this is done at a very early stage when neurons have not yet been produced and would be damaged by the process. In one surprising experiment, Evan Balaban and colleagues (Balaban et al., 1988) swapped quail for chick midbrain regions. Remarkably, the chicks with the quail midbrains continued to develop and hatched able to behave quite appropriately, except that they exhibited certain characteristic stereotypic movement patterns and calls of quails [unfortunately, both for them and for science, immune rejection quickly ensues, so these chimeric birds never get a chance to mature]. These brain inserts had matured, established extensive and appropriate interconnections with surrounding host brain structures, made appropriate functional links (even though the quail brain structures were much smaller than the corresponding chick brain structures would have been), and yet retained certain whole functional programs characteristic of the donor species. Clearly both highly integrated and complex cross-species interconnectivity develops as well as some level of species-specificity of function. These are not incompatible. In comparison with the pig xenograft procedures described above, however, one important difference is the implantation of whole intact brain regions (or rather their precursors). Thus, positional and growth cues within the midbrain region itself were appropriate to quail brains. The scaffolding that served as a source for selection and growth biases affecting neural connections within the midbrain were laid down by quail ontogeny. And it is likely that the connections necessary for producing these specific stereotypic bird behaviors are almost entirely contained within this segment of the brain (this has also been suggested by experiments involving electrical stimulation of and damage to structures located within the midbrain. A key factor in the evolution of such preprogrammed functions is almost certainly their embodiment within circuits that are localized within a relatively confined part of the brain, so that they are relatively insulated from systemic effects during development. They are modular in both structure and function. Highly constrained localized structures and modular function probably go hand in hand in this regard.

A characteristic feature of such modular functions is also that they are almost entirely automatic and fixed in their production. We would not normally call them intelligent behaviors, though they clearly may be components in higher order intelligent behavior patterns. It is important to recognize, however, that modularity of function also emerges out of the development of distributed systems, like that involved in skill learning. When we learn some motor skill, for example playing a scale with one hand on the piano, what begins as a process that involves a great deal of attention to sensory feedback and active inhibition of other movement tendencies, eventually becomes so automated that it can be performed as though activating an external mechanical device. Analysis of the distribution of brain activity before, during, and after learning simple mental skills clearly show that less and less of the brain is required to be involved as a skill becomes more and more automatic. Both the genetic evolution of such innately automated skills and the production of their facultative counterparts during life probably follow parallel paths of reduction of representation in the brain, but a great deal is yet unknown in both cases for us to be able to trace either process in detail.

How fly genes build mammal brains

The partial segmental independence of the development of connectivity within whole segments of the brain, such as is exemplified by the quail-chick transplants, is a reflection of a more general design strategy in embryology, that represents a middle level in the hierarchy of Darwinian processes that intercede between genes and connectional specification of the brain. It

is at this level, the level at which global features of brain organization are determined, that some of the crucial biases shaping later cellular and connectional selection processes are introduced, and so it is at this level where many characteristic species differences trace their ontogenetic origins. The discovery of the molecular genetic (DNA-RNA) basis for transmitting and decoding genetic information revolutionized biology more than any other discovery since Darwin. Recently, the discovery of a class of genes that regulate the large-scale patterns of regional differentiation of early embryos has revolutionized the study of development (see McGinnis and Kuziora, 1994, for a general review). Probably the most unexpected feature of these genes is that they not only do the major work of sculpting bodies and brains, but they are shared almost unchanged in species as divergent as flies and humans (Holland et al., 1992).

These genes share the common attribute of being able to regulate the activation of other genes because they include one or more base sequences that code for protein structures that bind directly to DNA. Many of these genes are called homeotic genes because they partition the embryo into segmental divisions that exhibit similar (homological) structures. Examples of segmentally homologous structures are found in the vertebral column and similar fore- and hind-limb structures. The brain and spinal cord initially develop from a neural tube that runs down the length of the embryo. Shortly after this tube forms (by an infolding of the embryo surface) it is partitioned by the localized expression of specific sets of homeotic genes into segmental regions that will eventually become the major brain and spinal cord divisions. What is remarkable about these genes is that almost identical genes are activated in the fly body during its development (and indeed many other groups of animals including worms to snails) in the same order and relative positions as in vertebrate embryos, presumably including humans. So the genes that determine early stages of human brain development also have close relatives involved in the early development of fly brains. This similarity verified a hypothesis offered by many comparative anatomists that most bilateral animals share a common logic of body plan. Until the discovery of homeotic genes, however, no one could have guessed just how conservative these pattern-generating mechanisms would turn out to be.

Though it is not quite clear how these genes interact in order to produce the patterning of expression over the whole embryo body, it is clear that it requires combinatorial interactions between these genes. For example, the earliest stages involve gene expression patterns that distinguish simple body axes (concentration patterns for different gene products that differ from top to bottom and from front to back) (e.g. see review by Boncinelli and Mallamacim, 1995). These patterns subsequently determine the patterns of expression of others in alternating bands of front to back stripes, and then, within this grid-like system, yet other genes are activated to further subdivide and distinguish these segments (McGinnis and Kuziora, 1994). This suggests that competitive interaction between families of similar regulatory genes vying to activate or inhibit one another, determine the pattern of when and where they each get expressed (Hoey and Levine, 1988). One of the important features of the expression of these genes in different parts of the embryo, is that they form highly discrete boundaries, which impose a kind of digital logic on an otherwise continuous sphere of cells. And this keeps the cell's progeny in line, so to speak, since there appear to be cell-cell adhesion properties expressed in conjunction with these gene products that cause cells within one segment to adhere and congregate but cells from adjacent segments to be excluded (Keynes and Krumlauf, 1992). This process enables a certain degree of independence of differentiation programs in nearby regions, yet the serially homologous organization guarantees spatial as well as molecular compatibility between segmental regions at the same time.

Though some homeotic genes continue to be expressed throughout development, most are activated at comparatively early stages in embryogenesis and are shut off shortly after, and almost certainly their roles as determinants of large scale morphology are played out early in the process. Correspondingly, organogenesis tends to be initiated quite early in embryogenesis and most of later development of the body is characterized by growth processes that enlarge these initial organ systems, but neither add to their numbers nor modify their basic spatial interrelationships. The reason for this probably has a lot to do with the limited distances across which competitive molecular interactions can produce regular patterns of expression. Diffusion rates of these macromolecules probably set an upper limit to the size of embryo that be partitioned by this means. As a result major divisions of the brain and body need to be determined at this small stage, and the rest of growth extrapolates from this starting point. This highly similar starting point for extrapolative growth probably contributes much of the regularity in allometric structural relationships among related species, even though not all structural divisions grow to the same extent.

Segmentation of the neural tube and growth of corresponding brain regions follows this logic. The expression domains of early regulatory genes in the head end of the neural tube identify regions that go on to become distinct sorts of neural tissues. These domains are initially laid down in a sort of front to back sequence of zones called neuromeres. In the brain stem and spinal cord these are simple segments identified by the sequential expression of a set of genes called Hox genes (named for the homeobox gene cluster in fruit flies where they were first identified), which are arranged in four paralogous (homologues in the same genome) linear clusters (see reviews by Kappen and Ruddle, 1993; McGinnis and Kuziora, 1994). The order of these genes on the chromosome and the order of their temporal and spatial expression from the base of the midbrain to the tail end of the neural tube is the same in flies and mammals (and probably all lineages in between). Most vertebrates differ by having 2 to 4 partially redundant clusters, the significance of which is not yet known.

Ahead of the midbrain-brainstem transition things get a little less linear. A number of other regulatory genes are expressed in a quasi-segmental pattern, obeying many of the same divisions of midbrain, diencephalon, and telencephalon (moving forward along the neural tube), but these are not arranged in as regular a pattern as in more caudal regions. Using a number of gene expression patterns and correlating them with morphological landmarks, Rubenstein, Puelles, and colleagues (see Rubenstein et al., 1994) have nevertheless identified a series of serially arranged prosomeres (neuromeres of the prosencephalon, which is the name for the undifferentiated forebrain at early stages), each of which is distinguished from adjacent ones by very distinct boundaries of gene expression. However, within the forebrain there seems to be many very distinctive gene expression divisions between dorsal and ventral regions, suggesting that one might need to additionally think of these as separate half-prosomeres at the very least. At present, there is no single logic that explains this topology and considerable controversy about whether a simple continuation of segmental organization is valid for this brain region, however, most agree that these unambiguous divisions mark boundaries between regions that will develop into distinctive brain regions with distinctive cell types and functions.

Two small families of regulatory genes do appear to follow a temporal and spatial expression pattern in the forebrain that is a sort of mirror image of Hox gene expression along the brain stem and spinal cord. They are identified as Emx and Otx (again named after fly counterparts-"empty spiracle" and "orthodentical"—that were discovered first in fly heads and were then used to fashion molecular probes to look for mammal homologues). The two

paralogues of each are expressed in a back to front, chinese box sort of pattern (one within the domain of the other) in the order Otx2>Otx1>Emx2>Emx1, with Otx2 essentially defining most of the forebrain ahead of the brain stem and Emx1 covering the smallest region, being confined only to dorsal telencephalon (which becomes the neocortex) (Simeone et al., 1992; Boncinelli et al., 1993).

Recent work with frogs suggests that the domain of Otx2 expression can be enlarged or reduced by either increasing its representation in all cells (by genetically inducing extra expression in all cells from an earlier time point) or inhibiting it with a differentiation factor called retinoic acid (reviewed by Boncinelli and Mallamaci, 1995). When its expression is augmented it correlates with an overdevelopment of the head end of the neural tube and when its expression is inhibited with a reduction of the head end. This, again, is consistent with a sort of competitive expression process, and additionally suggests a possible genetic basis for shifts in the segmental relationship between brain and body, such as distinguishes primates from most other mammals and distinguishes us from other primates (the subtleties of these growth differences cannot be pursued here, but see Deacon, 1991, 1995, and 1997, for discussions). To date there is little evidence that the neocortex is subdivided by different homeotic gene expression domains, which is consistent both with its relatively uniform cellular structure and its multi-potentiality during development. This should not be all that surprising from an evolutionary and developmental perspective, since only in more recent, and particularly terrestrial vertebrates has this region of the brain become particularly prominent. At the time in development that most homeotic genes are expressed, the entire telencephalon is one of the smallest divisions, perhaps composed of only two prosomeric regions, and so its few subdivisions are consistent in scale with those in other brain regions. The vast proportion of its disproportionate expansion occurs later in development in terrestrial vertebrates (and especially in birds and mammals), after these highly conserved genetic partitioning events have taken place.

The distinctions and divisions between these homeotic gene expression domains appear to play a major role in directing early generic axon growth. Early axonal path-finding tends to show abrupt growth effects that differ across neuromere boundaries and genetic or transplantation manipulations of the position of these early gene expression territories can produce corresponding redirection of axonal growth (see for examples Figdor and Stern, 1993; Kessel, 1993). It is probably the case that the regional differences in molecular activity in the brain provide the basis for crude target directed growth that gets axons to their destination in preparation for synaptic competition within it. Persistence of some of these homeotic guidance and specificity cues may be the source of the generic species-general growth patterns that guide appropriate axon growth in xenografted brains. Homeotic gene expression patterns may even play some role in introducing more subtle biases. For example, O'Leary and colleagues have recently shown that the retinotopic map patterning of projections from the eye to the tectum in chicks is influenced by the graded expression pattern of a gene called engrailed (also first found in flies). Retino-tectal axons grow in different directions in the tectum depending on the concentration of engrailed protein expression and on whether the axons originated from one sector of the retina or another. This may help explain why divisions of the cerebral cortex depend so much on Darwinian processes to be specified. Recall that this part of the brain is an expansion of a single homeotic segmental division, sharing common axon guidance cues and cell types throughout its tangential extent. The exception that proves this rule appears to be the fact that Otx gene expression also distinguishes deep cortical layers from superficial ones (Frantz et al., 1994). This is correlated with the fact that whereas different sectors of cortex are not

distinguished by different connection affinities, the different layers within each cortical region do have distinctive target specificities from one another that are shared in common with same layer cells in other regions. The deepest layer is layer 6 and its neurons only project to thalamic targets. Next, layer 5 neurons project to the most distant subcortical targets including striatum, midbrain, brain stem and spinal cord, but not thalamus. All superficial layer neurons project to cortex either locally or more distantly and only very few deep layer 3 neurons find their way to the nearest noncortical target, the striatum. So this exception provides further evidence for a conserved homeotic determination of initial connectivity.

In summary, the genetically specified tissue differences and axon target specificities of vertebrate brains all appear to be determined by a very conserved pattern of intercellular gene interactions at a very early point in embryogenesis. But although vertebrate brains all share most of the same major divisions and exhibit similar subdivisional architecture within them, there is nevertheless considerable variability among major classes. Only in mammals, for example, is there a thick many layered cerebral cortex. So it is not a totally conserved system. Very likely we will uncover homeotic gene expression differences that correspond to these brain structure differences between vertebrate classes, but at present there are too few cross-species comparisons to allow us even an educated guess as to what they might be, except to say that they probably involve the same highly conserved genes (or additional duplicates of them), perhaps shifted in expression domains or timing.

What makes human brains human?

On the basis of what else we know about human brains as compared to other species' brains, we can venture some reasonably educated guesses about what changes in ontogenetic processes caused us to deviate from more typical ape patterns. First, we can rule out a number of possibilities. There is no evidence of addition of novel parts. There are no extra neuromeric segments, no new kinds of brain structures, no special purpose devices plugged in, and almost certainly no new kinds of homeotic genes. Nevertheless, human brains have more of something making our brains large. Ours is a primate brain in every respect, except it is peculiarly large for our bodies. But our brains are not simply primate brains scaled up the way one might enlarge a photograph.

Comparisons of brains from large and small mammals in general, and large and small primates in particular, demonstrate some remarkably predictable cross-species size relationships among the parts of the brain. Plotting the sizes of two different brain regions with respect to one another in a spectrum of species of primates demonstrates a very orderly allometry (metric of differential growth) linking most together. Knowing the size of one of these major brain regions, it is possible to predict the size of most others with remarkable accuracy on the basis of the shared trend of that group, across a wide range of sizes. Not coincidentally, the regions that produce the most highly correlated trends turn out to roughly correspond to homeotic segmental divisions. This suggests that a common mitotic growth process may be applied to all, and that the trends reflect the way this scalar factor extrapolates the initial partitioning of brain regions.

But it is not a simple extrapolation because different parts scale according to slightly different trends. For example, a larger fraction of the brain in larger brained primates is forebrain (especially cerebral cortex) and cerebellum, and less of it is midbrain, brain stem and spinal cord (for a recent summary see Finlay and Darlington, 1995). This suggests that with increasing growth away from a common highly similar embryonic stage, slight differences in extrapolated growth have a greater and greater impact on relative proportions of brain structures

in different species. What determines the size of a given brain structure is of course the number of cell divisions that take place since homeotic processes have determined the ultimate cell fates for that region. This is largely a function of time spent in mitotic growth, since in mammals at least, there seems to be a relatively stable rate at which brains grow in the womb (e.g. Sacher and Staffeldt, 1975; Deacon, 1995; Deacon, 1997). One crude hint of how this might influence these proportional differences comes from comparing these trend differences with the timing of some of the underlying developmental changes. Those structures that exhibit a net positive allometry (i.e. tend to be proportionately larger in larger species) also tend to be the latest structures to mature, i.e. cerebral cortex and cerebellar cortex. As noted in the previous section, there is also a brainstem to telencephalon sequence of homeotic gene expression as well, so that the latest forebrain regions to mature are the latest to be carved up by the genes as well. Perhaps this consistent difference in reaching comparable maturational stages is responsible for a few additional cell divisions which, in larger species, get progressively extrapolated to greater proportional differences in the size of the corresponding brain structures. But whatever the reason, these allometric regularities once again indicate a remarkably conserved and predictable neural developmental process.

However, human brains don't fit the predictions in certain interesting ways. In each of the plots of comparative brain structure allometries, the human point is plotted for comparison to the other primate trends. Sometimes the human values are well within the range extrapolated from other primates (as in the relationship of thalamus to striatum), and yet in other cases the human values deviate considerably from the prediction (as in the relationship of thalamus to cerebral cortex). If the predictability of these patterns reflects conserved neuro-ontogenetic processes, then the human deviations indicate a break from the norm. Human brains aren't just scaled up primate brains, and they differ at the level of whole neuromeres. This implies that there may be some homeotic change that distinguishes human brain development from all other primate brain development. The Otx and Emx genes offer relevant candidates, since the structures that are disproportionately larger in human brains are initially defined by these genes' expression domains. What the change might be we have little hope of predicting at this time, except that it probably involves these or related gene expression events.

We can, however, predict some of the consequences from what we already know about the Darwinian nature of brain development. The shift in proportions in major brain regions should also affect the axonal competition for connections throughout the brain, just as would the addition of a large number of additional sensory receptors or extra limbs. It is the neurological analogue of introducing a major new resource into an ecosystem that preferentially benefits one species. The result is that all selection relationships are altered to some degree. Unlike the blind mole rat or the frog with an extra limb, this change is centrally located and so will have its major influence of connectional development from the inside out. I have argued elsewhere (Deacon, 1990; 1995; and 1997) that it would tend to influence sensory systems relatively little since they would be strongly constrained by their peripheral inputs, but that it would both increase cortical output control over numerous subcortical systems, including motor systems involved in laryngeal and mouth movements. It would also alter the relationships between cortical areas since most cortical connections are with other cortical areas anyway.

These are clues to exactly what differences in neural function were selectively favored in our ancestors. Was it just increased intelligence in some generic sense? The appearance of significant rearrangement of neural proportions and the ramifications this has for connectional reorganization suggests something quite different. It suggests that there has been a change in

another level of Darwinian processes: a cognitive level. It suggests that the balance of neural computations has been biased so that some systems within the brain have a significant competitive advantage in determining the course of the outcome of moment-by-moment neural activity patterns than others. From a Darwinian perspective, it represents a major innate shift in cognitive selection conditions that will favor patterns of neural activity produced in the more expanded brain regions over those from the less expanded ones: a shift in the 'ecology of mind' (to borrow a phrase from Gregory Bateson). Certainly our one most distinctive cognitive faculty, language, reflects such a shift, and not just a change in mental capacity or the implementation of a set of inborn grammar-decoding instructions. The puzzle is figuring out how the two are related (though this cannot be pursued here, I offer a detailed alternative theory of language evolution, acquisition, and neural processing that is based on this Darwinian analysis in my recent book *The Symbolic Species* (1997)).

Evolution turned inside-out

In every generation, it seems, the most sophisticated technologies are appropriated to serve as metaphoric models for the mysterious working of brain processes. Within a century our models have shifted from hydraulic systems to telephone switchboards, and now, to digital computers (which have had the reverse honor of being attributed brain functions like memory and intelligence). Even the computer hardware/firmware/software distinction has found its way into the metaphors of mental processes, as an analogue to neural connections/innate abilities/learned abilities, though any similarities are extremely superficial. The algorithmic and computer metaphors, like all good analogies, have nonetheless sharpened the precision of models of mental processes. One cannot only be explicit about predictions of these models, they can be implemented in computer simulations of mental computing to see how they behave in a variety of contexts. So these models are only troublesome if they are over-interpreted, if one mistakes the map for the territory. A few decades ago it was legitimate to withhold judgment about whether or not the processes of neural representation resembled algorithmic processes, i.e. explicit symbolic codes and subroutines that were explicitly stored, retrieved, and executed to run sensory analyses or motor movements in much the same way as machine instructions determine a specific sequence of signal processing transformations or revolutions of a stepper motor controlling a robotic arm. It made sense to wonder if innately predisposed response patterns were encoded and stored in the genome to be reloaded into the fetal neural memory banks, like uploading the operating system into RAM at machine startup. Thus, an innate universal grammar, specified as a set of rules for making certain classes of symbol transformations, could appear as a plausible model of human language acquisition predispositions.

But as we now begin to understand the details of the genes to brain translation we are faced with a very different sort of logic. If instead we take as our metaphors of mind the biological processes that underlie mental processes and the logic that goes into the construction of brains, a very different set of predictions about the logic of mental processes emerges, one that is not well modeled by computer metaphors precisely because computers and their software are designed.

Psychologists have long been comfortable (some would say too comfortable) with comparisons of high level mental operations, such as learning, with evolutionary processes. For example, superficially, trial and error learning processes resemble evolutionary processes. Alternative responses are like competing organisms, variations in response tendencies are like

mutations, and pleasurable or aversive outcomes are like natural selection which determines which responses are most likely to be repeated in the future. The analogy of learning to natural selection has been recognized almost from the time Darwin's ideas became public. Indeed, even before Darwin, the analogy was implicit in the Lamarckian conception of evolution, and writers like Herbert Spencer made it the core of his theory of Psychology. But also, because of this similarity, the role of learning has been a source of confusion in evolutionary theory, as Spencer and Lamarck also demonstrated. Our guesses and behavioral "trials" are seldom completely random and the responses we retain are seldom those optimally selected with respect to reinforcement contingencies, but to some extent the same analogies could apply to all levels of evolutionary processes. Prior biases can play a significant constraining role in determining what changes are more or less likely. Learning, then, adds an additional layer of evolution-like processes to the hierarchy. And producing this ability is one of the major reasons that brains evolved. But what is going on in the brain to support this?

A number of scientists and philosophers have turned to Darwinian processes for clues to try to explain this middle level of brain functions. The biologist Gerald Edelman (1987), the neuropsychologist Michael Gazzaniga (1985), the neurologist Jason Brown (1979), the physiologist William Calvin (1987; 1989), the philosopher of mind Daniel Dennett (1991), and the evolutionary theorist Henry Plotkin (1994), are just a few prominent thinkers who proposed variants on Darwinian theories of learning processes. In general, some approach along these lines must be the appropriate one. I will not attempt a survey of the similarities and differences, or strengths and weaknesses of these various theories here, and there are many important differences that will require empirical modeling and testing to ultimately sort out. Rather, I want to end by offering some general comments on how the underlying Darwinian processes should help us think about these alternative models, for these models too can be stopped short of a thorough-going Darwinian analysis the same way that the standard "design by natural selection" models have: by insinuation design logic somewhere in the middle without noticing it.

Consider the fact that patterns of neural activity that determines which neural circuits survive and which are eliminated in development. The sculpting of the connectional architecture of the brain during its development is thus a direct reflection of the underlying logic of neural information processing. And this process, we have seen, is a Darwinian process. So we may gain some hints by looking at what selectively favors the persistence of certain patterns of connection over others during development. The answer, as noted above, appears to be the degree of coherence of the signals carried by competing axons as they converge at each neuron. Dyssynchrony is selected against and axons transmitting signals that are least like the collective mean pattern will tend to be eliminated. Each neuron is essentially tallying votes and rewarding those who voted for the winning pattern. This offers two important clues concerning the global logic of neural signal processing. Correlated temporal patterning is likely to be important in selecting which signals are most likely to persist and spread through the nervous system, and different temporal patterns of activity will likely be in competition with one another for representation in succeeding moments and for being passed to other sites within the brain.

The likely importance of correlated temporal patterning is not news (much is based on the pioneering work of Donald Hebb; see for example Hebb, 1949). Since the discovery of electrical activity emanating from the cerebral cortex at characteristic frequencies (as measure by electroencephalography, EEG), neuroscientists have felt that some aspect of cyclic signal processing might be critical to organizing global brain functions. Gerald Edelman (1987) and Francis Crick (Crick and Koch, 1990) are two prominent Nobel lauriets who have suggested

alternative theories in which the establishment of temporal synchrony between widely separated brain structures is the key step in recruiting them to contribute to a common cognitive task. The production of an integrated behavioral response inevitably requires the concerted participation of a wide range of brain structures, and there is considerable evidence that alternative response tendencies arise in different brain regions, compete to recruit correlated activity in other brain regions, and ultimately the patterns that have recruited the most correlated activity and displaced the alternatives control behavior. This seems particularly evident in the motor system where the coordinated activity of large sectors of motor cortex are necessary to initiate directed limb movement. But where do such patterns originate in the first place? Are there specific neurons that are the drivers? Probably the key to this prime mover problem can be found in the intrinsic noisiness of neural signal production itself. Neurons are likely unreliable signal transducers and due to intrinsic metabolic and molecular fluctuations are to a significant extent spontaneously active, irrespective of input. This is a source of intrinsic noise. Rather than being a problem, however, this spontaneous noisiness may be the source of pattern generation, or rather that from which neural selection processes can extract adaptive patterns. The selection biases are produced by extrinsic signals supplied by the senses, intrinsic biases supplied by homeostatic mechanisms and the architecture of connectivity, both as determined by development and by the way memory traces bias the connections between patterns. There need be no prime mover in evolutionary processes, no specific instigator of a given response, no localized organizing initiating center, only sources of variety and biases that select from it.

Though this account of the global logic of neural information processing is highly speculative and vague, as all current theories of global brain function must be, it offers a model for understanding brain functions that is consistent with the overall dynamic of organism evolution and development in ways that approaches from "design" are not. Though we may be able to develop algorithmic and deductive models that capture some superficial aspects of our cognitive processes, they are unlikely to have any deep isomorphism with what really goes on inside brains, since this is inevitably statistical and Darwinian in character at almost every level.

So evolution is the origin of intelligence in both senses. Over billions of years, life has evolved to be more capable of internally representing ever more complex multileveled environmental relationships and producing adaptive responses to the predicaments they pose. It has done so spontaneously, by evolving means of embedding evolutionary processes within other evolutionary processes to produce complex multilevel organisms with greatly amplified adaptive information generating capabilities. I suspect that the most serious barrier to developing a full explanation of the evolution of intelligence, complex behavioral abilities, and mechanisms underlying conscious intentionality has been our failure to understand organism function from a fully Darwinian perspective. We have been ensnared by the reasonableness of engineering and design metaphors, and use them as short cuts to avoid a full evolutionary accounting of the many levels of ontogenetic, physiological, and social mechanisms involved in the generation of adaptive complexity. Following this design metaphor, we have been tempted to think of brain evolution as the addition of modules, brain development as design from a genetic plan, and cognition as a collection of algorithmic processes.

The key insight that leads beyond these theoretical cul-de-sacs is the realization that the evolutionary processes that have produced our bodies and minds are operating at all levels. Though genetic information provides the outside constraints within which all other organismic information processes must be encapsulated, it only accounts for a very minute fraction of the total information that must be generated to build an animal body, and an even smaller fraction of

what must be generated to specify the ongoing behavioral adaptations that an animal produces during its lifetime. This information must be spontaneously generated during development and must be generated moment by moment in behavior and cognition. The sources of this new information are distributed in invariant relationships that already exist, and are internalized by higher order Darwinian processes.

Some of the most compelling analyses of high level cognitive processes have been presented in terms of functional design, especially those thought to be strongly dependent on unlearned information. A well known example is the cognitive basis for grammatical analysis and language acquisition characterized by theories of an innate universal grammar (for a current interpretation of this position see the chapter by Steven Pinker, as well as his recent book The Language Instinct). Currently, the wonderfully suggestive title "Evolutionary Psychology," ["Darwinian Psychology" in original] has become synonymous with a whole class of similar accounts of behavioral and cognitive predispositions (for discussion see the chapter in this volume by Tooby and Cosmides). In summary, these views share in common the notion that the human mind can be modeled as a large collection of algorithm-like information processors, or computational modules that have been designed by evolution to function in highly selective information processing domains, and which essentially carry out their operations in isolation of other modules with which they share inputs and outputs. This is an explicit view of intelligence as designed, albeit in piecemeal fashion, with natural selection as the designer. Though the Darwinian physiological approach I have been outlining does not exclude the possibility that highly modular sensori-motor or cognitive predispositions might be present from birth, it does suggest that modeling these predispositions in terms of algorithms, deductive rule systems, or lists of outputs will provide misleading heuristics for understanding the evolutionary processes and information processes that underlie them. Moreover, the Darwinian perspective suggests that such a modular design strategy is the exception and not the rule. If there is an efficient way of achieving the same adaptive end without the need for prior specification of the information, it will be used instead. The molecular and cellular processes that produce brains operate according to a logic very unlike that of a computer programmer devising a series of operations to serve a specified purpose, and what they produce is as a result very unlike a computer algorithm. Even if we argue that a given neural operation is strongly prespecified by intense selection, the challenge is to discover how the design metaphors we use to describe its operation can be realized in terms of a Darwinian ontogenetic process. But it may be better to recast these models in Darwinian terms from the start, abandoning the design metaphor entirely.

In general, the evolution of behavioral information processing has not produced brains that are analogues of general-purpose computers, nor has it produced brains that are merely collections of special-purpose minicomputers each designed to fit a particular evolutionary task. Though it has produced approximations to both of these design strategies in some of its different parts, these are extreme special cases in neural development (at least in mammal brains, though not necessarily in small invertebrate brains). For the most part, what is behind the intelligent design and function of brains—even in these special cases—is the evolution of subsidiary Darwinian process embedded one in the other that are thereby able to vastly amplify the scope of the adaptive "search" process and relieve the genome of this impossible task. Brain functions are the most elaborated expressions of such a multilevel hierarchy of Darwinian processes.

The experience of intelligence

What insights does this Darwinian approach to organism design have to offer us in approaching that last terra incognita of neuroscience: consciousness and experience of generating intelligent thought and action? Philosophers have long struggled with the problem of making sense of the subjective experience of consciousness in the context of the deterministic laws of the chemistry and physics that must underlie mental processes. We all have an undeniable experience of being sources of action and consideration, of deciding what to do, of resisting compulsion, and of reasoning about the best course of action before we act accordingly. Robots in which present global states are entirely prefigured in prior global states of its mechanism, or whose computational behaviors are dictated by the execution of the next line in some computer program code, should experience neither conflicted nor indeterminate states. Nevertheless, except for processes at the subatomic level (where a few desperate theorists have looked for a loophole), physical events are remarkably well determined by prior conditions. This is the source of that age old conundrum about whether we have free will or are blindly determined in our behaviors—a problem that more than once has found its way into courtroom questions of sanity and responsibility. If consciousness is a function of the actions of matter and energy at all levels, then how could we really be the spontaneous originators of new actions and thoughts? How could we fail to be prefigured in what came before? Aren't we in fact mere mechanisms, chemical machines performing determined computations?

It should now be clear that not all mechanisms are mechanisms in the way we tend to think of clockworks: boringly mechanical and predictable. Organisms are not just somewhat messy and unpredictable on the surface. They are deeply indeterminate systems through and through. They are the opposite of clockworks in this regard, and are level upon level of unspecified statistical evolution-like process embedded within one another. Mental activity is simply the most embedded level, and is most functionally removed from simple determinative relationships by many levels in between. Neural circuits are the asymptotic extension of Darwinian neural developmental processes driven by the signal processing logic that neurons employ. But what is this logic? One hint is the developmental process itself. Apparently the generation of the complex patterns of connectivity that constitute our brains are not strongly predetermined during development, but come into being in an evolution-like process driven by the flow of signals within this forming network. The outgrowth of neural processes spontaneously generates an immense variety of patterns that are sculpted and evolved in microcosm under selection pressures derived from this signal processing. By this means brains generate circuit complexity to match the complexity of the signals passing through them. But there is no reason to imagine that once relatively stable adult neural circuits are formed that the signal processing logic that formed them should suddenly take on an entirely different character. If we can use the logic of neural circuit formation in embryogenesis as an index of the logic that produced it, we inevitably must conclude that neural signal processing is fundamentally driven by some kind of evolutionary logic.

This is also the inevitable conclusion we must draw if we just consider the nature of the components out of which brains are built. Neuronal signal processing is supported by molecular metabolism and is rife with local stochastic effects and source for considerable spontaneous neural firing. It generally appears that only a small fraction of the activity of an individual neuron can be accounted for by so-called trigger features that drive its more intense "spiking" behavior (though it is also possible to explain this as secondary responding to diverse other inputs, spontaneous firing almost certainly accounts for some significant fraction of this; how

much is difficult to tell). But if even a small fraction of the activity of individual neurons is from spontaneous firing the effects will be overwhelming. This is because it would mean that the vast vast majority of signals in the brain arise intrinsically, not extrinsically, since vastly more neurons are buried within the CNS than transduce patterns from external sources. The brain is anything but a passive responder to input. It is almost as though input signals are nearly superfluous compared to self-generated signals. Moreover, much of the brain's circuitry, and particularly that within the cerebral cortex, is linked together by excitatory synaptic connections. This will tend to amplify any intrinsic spontaneously generated "noisy" signals. This "design logic" is just the opposite of what would be necessary to produce well-behaved computation. Certainly brains are not computers in anything like the usual sense of that concept. The only hope in such an intrinsically chaos-amplifying system, is to somehow use the noise against itself, so to speak. Evolution works by just such a logic. Competing randomly perturbed variants, generated in massive numbers are pitted against one another under highly specific conditions that mediate, via selective inhibition or enhancement of their persistence and replication, which will predominate and serve as the seeds for future variants. Thus, it seems that the properties of brains and the traces of signal processing left in the wake of their development point to an evolutionary logic at work at this level of organism physiology as well.

This makes good sense from an adaptive functional perspective, because the problem of generating mental actions is simply another version of the problem of spontaneously generating novel complex adaptive responses in general in a context where the form of the adaptation is not predetermined and the combintorial details are far too complex to be managed individually. Our brains keep up with the massively information-rich input-output context presented to them by continuously generating new adaptive complexity. Like embryogenesis and the production of immune cells, brains must generate useful information from scratch, in order to fit in with and take advantage of a novel environment even as it presents itself. And so in order to be prepared for the widest range of alternative possibilities and sample this range effectively, the initial generation of variety must be uncoupled as much as possible from regularities in the input with respect to which it is selected. But of all the problems of generating adaptively complex patterns that we have considered, the problem of generating intelligent behavior is one of the most demanding. Faced with the onslaught of an unimaginably diverse, unpredictable, and massive input, demanding an adaptive response on line at every moment, only an evolution-like process could be capable of generating a sufficient level of appropriate responses.

Perceptual processes are not merely classification-mapping algorithms and behavior is not merely the activation of alternative outputs in response to specific perceptual signals. Both involve competitive interactions between alternative neural activity patterns each competing for wider representation in multiple brain regions. Perceptions and actions are the expression of the temporary winners' widespread representation. Ongoing spontaneous nonspecific signal production is inevitable in a brain full of noisy processing elements, and biases affecting which signal patterns are most likely to out-compete which others in specific brain regions are both accumulated during development in the form of connection patterns and during later life in the form of differing synaptic strengths due to learning. On top of these relatively stable biases, externally originating signals from the sensory organs add a constantly changing adaptive challenge at a micro as well as a macro level of brain function. Spontaneously generated, diffusely produced, and nonspecific neural activity provides the raw materials for these sensory and intrinsic influences to selectively preserve and amplify patterns by virtue of their self-reinforcing effects, post hoc.

How does this relate to the subjective experience of being such a process? Our self-experience is one of originating, moment by moment, novel detailed and appropriate representations of an ever-changing environment and spontaneously generating adaptive responses to them without many specific detailed predetermined alternatives. The life of the mind, so to speak, does not have a mechanical character, except in the form of compulsions to act. We easily distinguish those times when we act on impulse or unconsciously perform some rote response, from those times when we act intelligently and rationally. We experience a clear difference between such compelled behaviors as laughter, sobbing, and sneezing, which produce actions irrespective of reflection, and the processes used to inhibit their production. It often requires something analogous to physical effort in order to resist these compulsions. And similar competitive mechanisms are part of our experience at the lower levels of all neural systems from the tendency to become distracted by interesting sensory stimuli to the difficulty of maintaining traces in working memory long enough to complete a complicated mental computation using them.

In a deterministic device, the microscopic noisiness or imprecision of the operations of the parts is a problem to be overcome. Different scales of deterministic processes are deliberately kept apart. The point of good engineering design is to minimize the disturbing influences of the low level messiness of the device on its global performance, thus keeping its operation descriptively simple. We want our watches and motors to operate the same way irrespective of changes in friction, expansion or contraction due to temperature change, and changes in shape as a result of wear. In digital processing components, for example, it is essential to design the semiconductor elements so that although the electric signal imposed upon it is continuously fluctuating, these unavoidable small variations in the voltage are all below the level of the voltage differential required to change the state of the material from non-conductive to conductive. In other words, signals and noise are categorically distinguished, as are signals and large-scale external perturbations. But this is precisely the opposite for a Darwinian process. Low level noise is the source from which signals are selected and high level perturbations provide the bias that can selectively amplify or damp micro-variations to reach or fail to reach some threshold strength, the value of which is itself stochastically varying. One can't have it both ways. Strategies for excluding the effects of noise in order to produce determinate predictability make it impossible to use noise as a source of novel responses to the unpredictable. Organisms have evolved to take advantage of the power of their own intrinsic noise via Darwinian information processing strategies at all levels of organism design and function where they are under pressure to generate adaptive information to overcome the inadequate information capacity of genomic inheritance. They must, as a result, increasingly rely on selection processes at progressively higher levels of adaptive organization to provide it, because these are the only processes capable of generating novel adaptive complexity autonomously and without pre-specification.

In summary, it seems indisputable that awareness is a property that physical mechanisms can exhibit, However, a clockwork mechanism, designed so that its processes are minimally affected by the micro-fluctuations of properties of its constituent parts would not be capable of experiencing the world, irrespective of the level of complexity of its workings. Simply mapping inputs onto outputs, no matter how conditional and complicated the mapping algorithm, cannot have awareness as a correlate. There is no complexity threshold separating mindless mechanism from aware mechanism. There is, rather, a fundamental difference in architecture and in the mode by which energetic processes evolve into signal processes that distinguishes these two

kinds of mechanisms. Mechanisms whose functional relations to possible classes of input signals are pre-specified by design are ultimately informationally static, even though their operations may be complex and conditional. Like deductive proofs, whose results are implicit before they are derived, determinate machines, whose subsequent states can be predicted from prior states and current inputs, are closed to information--windowless monads to use Leibnitz' colorful phrase. Precisely the opposite is true of evolution-like processes, which are formed around a sort of fundamental openness, incompleteness, and vagueness of determination. In a true evolutionary process (as opposed to incompletely open simulations such as many genetic algorithms), it is never possible to predict outcomes from prior states, except in a very general and global form. A process that is capable of amplifying even the weakest of spontaneous perturbations to become major influences on subsequent states is deeply indeterminate in the real world. This kind of organization would be a recipe for chaotic self-undermining behavior were it not for the taming influence of selection, in which aspects of this chaotic activity can be pitted against itself in the context of adaptive biases. Still the essential dominance of chaos expresses itself in our experience of mental processes which are mostly messy, incomplete, nebulous, and indeterminate in character. Constrained chaos.

Awareness implies not only responsiveness to the flow of novel information from outside, but also the comparison of this flood of input to an equally incessant flow of internally generated novel self-information. Subjective experience is born in that encounter between the outflow of spontaneous signal generation and the inflow of signals from the world. The experience of being an intelligent organism, is I believe the experience of being an evolutionary process in action. It is the experience of becoming an adaptation, from the perspective of the generative element itself, i.e. the noise. We are the chaos, the self-sustaining entropy of life. But neural chaos would only be chaos except for the fact that it incessantly encounters itself, and is forced to be in adaptive competition with itself (or parts of itself) within the constraints of the neural nets that contain it and with respect to the extrinsic signals that selectively bias it.

So what does this imply about the subjective experience that we refer to as awareness or consciousness or sentience or what have you? First, it implies that it is not just some epiphenomenal illusion, but a necessary constituent of neural information processes. Second, it implies that the sense of self-origination that characterizes this experience is not merely incidental to some more fundamental mindless "neural computations" that comprise the working of the brain, but can't be dissociated from this neural activity. Third, though my perspective is in opposition to the tenets of much of contemporary computational reductionism, it does not imply that minds depend on some nonmaterial essence or some ineffable and mysterious mind stuff that is forever out of the reach of scientific inquiry. The nature of subjective experience is not beyond our comprehension, and will not forever be beyond our capacity to create from scratch in mechanisms of human design. Though it may seem counterintuitive at first to think in these terms, I believe that understanding the logic of how this is possible is a bit like entertaining the Zen *koan* posed at the outset of this essay. What is it like to be evolution?

Information can be created and destroyed. This is because, unike the stuff of the physical world, information is a property of relationships not of things in themselves. Patterns with respect to patterns which can be destroyed (or rather irreversibly complicated) by entropic processes, and which can also be brought into being (or rather uncomplicated or de-convolved) by that subtle variant of entropy called evolution. Both biological evolution and mental processes are information creating processes, and not surprisingly we identify ourselves with the locus of this created information in both senses, as living experiencing entities, and we tend to

meld these two aspects into one and the same notion in our common sense psychology. This is nowhere better exemplified than in the popular modern day Pinocchio movie of the 1980s, called Short Circuit, in which a robot attempts to describe his miraculous transformation (via lightning bolt) into a conscious being with the exclamation "I'm alive!"

To generate new information is to be aware, even if only in the limited sense in which all living organisms possess some measure of it. This underlies the deep analogy between life and mind. While there may be no "homunculus" outside the system that directs the flow of information through our nervous systems, there is a spontaneous non-predetermined self-organized logic to how we operate. The homunculus fallacy is an artifact of considering organisms from an engineering "design" perspective. But organisms have not been *designed* and are neither functionally organized nor produced the way designed artifacts are. The logic of organisms is an evolutionary logic through and through, and not just antecedent to an organism's life. The evolutionary logic is exemplified in the way organisms assemble themselves at every stage and level of the life process, all the way up, from embryogenesis to cognition. We feel as though we are the originators of our novel behavioral responses and at the origin of a spontaneous fountain of new information erupting into the world, because we are. This core subjective experience of self-origination is neither illusory nor misleading. We are indeed what our self-experience suggests: evolution in action.

References

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

Boncinelli, E., Gulisano, M. and Broccoli, V. (1993) Emx and Otx homeobox genes in the developing mouse brain. Journal of Neurobiology 24, pp. 1356-1366.

Boncinelli, E. and Mallamaci, A. (1995) Homeobox genes in vertebrate gastrulation. *Current Opinion in Genetics and Development 5*, pp. 619-627.

Brown, Jason W. (1979) Language Representation in the Brain. In H. Steklis and M. Raliegh (eds.), *Neurobiology of Social Communication among Primates*. New York, Academic Press, pp. 133-195.

Calvin, W. (1987) The brain as a Darwin Machine. *Nature* 330, pp. 33-34.

Calvin, W. (1989) A global brain theory. *Science* 240, pp. 1802-1803.

Cowan, W., Fawcett, J., O'Leary, D. and Stanfield, B. (1984) Regressive events in neurogenesis. *Science 255*, pp. 1258-1265.

Crick, F. and Koch, C. (1990) Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences 2*, pp. 263-275.

Dawkins, R. (1976) *The Selfish Gene*. Oxford: Oxford University Press.

Dawkins, R. (1986) *The Blind Watchmaker*. New York: W. W. Norton.

Deacon, T. (1990) Rethinking Mammalian brain evolution. *American Zoologist 30*, pp. 629-705.

Deacon, T. (1995) On telling growth from parcellation in brain evolution. In E. Alleva, A. Fasolo, H.-P. Lipp, L. Nadel, and L. Ricceri (eds.), *Behavioral Brain Research in Naturalistic and Semi-naturalistic Settings*, Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 37-62.

Deacon, T. (1997) *The Symbolic Species: Coevolution of Language and the Brain*. New York: W. W. Norton.

Deacon, T., Pakzaban, P., Burns, L., Dinsmore, J. and Isacson, O. (1994) Cytoarchitectonic development, axon-glia relationships and long distance axon growth of porcine striatal xenografts in rats. *Experimental Neurology 130*, pp. 151-167.

Dennett, D. (1991) *Consciousness Explained*. Boston: Little Brown.

Doron, N. and Wollberg, Z. (1994) Cross-modal neuroplasticity in the blind mole rat Spalax Ehrenbergi: a WGA-HRP tracing study. *Neuroreport 5*, pp. 2697-2701.

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

Figdor, M. and Stern, C. (1993) Segmental organization of embryinic diencephalon. *Nature 363*, pp. 630-634.

Finlay, B. and Darlington, R. (1995) Linked regularities in the development and evolution of mammalian brains. *Science 268*, pp. 1578-84.

Finlay, B., Wikler, K. and Sengelaub, D. (1987) Regressive events in brain development and scenarios for vertebrate brain evolution. *Brain Behavior & Evolution 30*, pp. 102-117.

Frantz, G., Weimann, J., Levin, M. and McConnell, S. (1994) Otx1 and Otx2 define layers and regions in developing cerebral cortex and cerebellum. *Journal of Neuroscience 14*, pp. 5725-5740.

Frost, D. O. and Metin, C. (1985) Induction of functional retinal projections to the somatosensory system. *Nature 317*, pp. 162.

Gazzaniga, M. (1985) The Social Brain: Discovering the Networks of the Mind. New York: Basic Books.

Gould, S. J. and Lewontin, R. C. (1979) The spandrels of of San Marco and the Panglossian program: A critique of the adaptationist program. *Proceedings of the Royal Society of London 205*, pp. 281-288.

Hebb, Donald (1949) *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.

Hoey, T and Levine, M. (1988) Divergent homeobox proteins recognize similar DNA sequences in Drosophila. *Nature 332*, pp. 858-861.

Holland, P., Ingham, P. and Krauss, S. (1992) Development and evolution. Mice and flies head to head. *Nature 358*, pp. 627-628.

Isacson, O., Deacon, T., Pakzaban, P., Galpern, W., Dinsmore, and Burns, L. (1995) Transplanted xenogeneic neural cells in neurodegenerative disease models exhibit remarkable axonal target specificity and distinct growth patterns of glial and axonal fibres. *Nature Medicine 1*, pp. 11898-1194.

Isacson O, and Deacon T. (1996) Specific axon guidance factors persist in the adult brain as demonstrated by pig neuroblasts transplanted to the rat. *Neuroscience 75*, pp. 827-837

Jackendoff, R. (1992) Languages of the Mind. Cambridge, MA: MIT Press.

Kappen, C. and Ruddle, F. (1993) Evolution of a regulatory gene family: HOM/HOX genes. *Current Opinion in Genetics and Development 3*, pp. 931-938.

Katz, J. and Lasek, R. (1983) Evolution of the nervous system: role of ontogenetic mechanisms in the evolution of matching populations. *Proceedings of the National Acadamy of Sciences, USA 75*, pp. 1349-1352.

Kauffman, S. (1992) *Origins of Order: Self-Organization and Selection in Evolution*. Oxford: Oxford University Press.

Kessel, M. (1993) Reversal of axonal pathways from rhombomere 3 correlates with extra Hox expression domains. *Neuron 10*, pp. 379-393.

Keynes, R. and Krumlauf, R. (1994) Hox genes and regionalization of the nervous system. Annual *Review of Neuroscience 17*, pp. 109-132.

Killackey, H., Chiaia, N., Bennett-Clarke, C., Eck, M. and Rhoades, R. (1994) Peripheral influences on the size and organization of somatotopic representations in the fetal rat cortex. Journal of Neuroscience 14, pp. 1496-1506.

Law, M. and Constantine-Paton, M. (1981) Anatomy and physiology of experimentally induced striped tecta. *Journal of Neuroscience 1*, pp. 741-759.

McGinnis, W. and Kuziora, M. (1994) The molecular architects of body design. *Scientific American 270*, pp. 58-66

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

O'Leary, D. and Stanfield, B. (1989) Selective elimination of axons extended by developing cortical neurons is dependent on regional locale experiments utilizing fetal cortical transplants. *Journal of Neuroscience 9*, pp. 2230-2246.

O'Leary, D. (1992) Development of connectional diversity and specificity in the mammalian brain by the pruning of collateral projections. *Current Opinions in Neurobiology 2*, pp. 70-77.

Paley, W. (1802*) Natural Theology: or, Evidences of the Existence and Attributes of the Deity, Collected from the Appearances of Nature*. London: E. Paulder.

Pinker, Steven (1994) *The Language Instinct: How the Mind Creates Language*. New York: William Morrow & Co.

Plotkin, H. (1994) *Darwin Machines*. Oxford: Oxford University Press.

Purves, D. (1988) *Body and Brain. A Trophic Theory of Neural Connections*. Cambridge, MA: Harvard University Press.

Purves, D. and Lichtman, J. (1980) Elimination of synapses in the developing nervous system. *Science 210*, pp. 153-157.

Roth, G., Nishikawa, K. C., Naujoks-Manteuffel, C., Schmidt A. and Wake, D. B. (1993) Paedomorphosis and simplification in the nervous system of salamanders. *Brain, Behavior & Evolution 42*, pp. 137-70.

Rubenstein, J., Martinez, S. Shinmamura, K, and Puelles, L. (1994) The embryonic vertebrate forebrain: the prosomeric model. *Science 266*, 578-580.

Sacher, G. and Staffeldt, E. (1974). Relation of gestation time to brainweight for placental mammals: Implications for the theory of vertebrate growth. *American Naturalist 108*, pp. 593-615.

Simeone, A., Acampora, D., Gulisano, M., Stornaiuolo, A. and Boncinelli, E. (1992) Nested expression domains of four homeobox genes in developing rostral brain. *Nature 358*, pp. 687-690.

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

Sur, M., Garraghty, P. and Roe, A. (1988) Experimentally induced visual projections into auditory thalamus and cortex. Science 242, pp. 1437-1441.

Wilczynski, W. (1984) Central neural systems subserving a homoplasous periphery. *American Zoologist 24*, pp. 755-763.

Yamamoto, N., Yamada, K., Kurotani, K. and Toyama, K. (1992) Laminar specificity of extrinsic cortical connections studied in coculture preparations. *Neuron 9*, pp. 217-288.