From Biology To Consciousness To Morality

Ursula Goodenough and Terrence W. Deacon*

ABSTRACT Key Words: biology, consciousness, morality, emergence, brains, symbolic language, culture, moral ideals, moral motivation, virtue
Social animals are provisioned with pro-social orientations that transcend self-interest. Morality, as used here, describes human versions of such orientations. We explore the evolutionary antecedents of morality in the context of emergentism, giving considerable attention to the biological traits that undergird emergent human forms of mind. We suggest that our moral frames of mind emerge from our primate pro-social capacities, transfigured and valenced by our symbolic languages, cultures, and religions

Introduction

One of us has recently offered a strong claim: "Biologically we are just another ape; mentally we are a whole new phylum of organism" (1).

So how did this come to be? How does our apparently novel mentality, and its attendant sense of self, relate to our evolutionary heritage? We offer here some perspectives on these questions, with a particular focus on the dynamics of our moral sensibilities.

We begin with the concept of emergence. We point out that biological emergence is undergirded by semiotic (encoding) systems, and describe how such systems are manifested in single-celled and multicellular organisms, particularly as they generate cellular awareness and, in animals, brain-based awareness. We then note that a unique semiotic system — symbolic language — has evolved in the hominid lineage, and offer a scenario for the unfolding of that evolutionary process. We conclude by proposing that a core feature of our mentality is our ability to access and experience primate states of mind, and that our moral capacities are rooted in this dynamic. Whereas human transcendence is commonly configured as a "from-to" trajectory towards the beyond, we suggest that much of human transcendence entails a circling back to the "from" dimension and transfiguring it with our symbolic minds.

Emergence

The explosion of interest in emergentism ("something more from nothing but") among both natural scientists and philosophers has not surprisingly generated considerable ambiguity in what is meant by the term emergence. One of us (2) has therefore offered an inventory of emergent phenomena in the natural world, proposing that emergence takes three forms:

- First-Order Emergence: Properties emerge as a consequence of shape interactions. Example: The interaction of water molecules (nothing but) generates a new property, surface tension (something more).
- Second-Order Emergence: Properties emerge as a consequence of shape interactions played out over

time, where what happens next is highly influenced by what has happened before. Example: The formation of a snowflake, where initial and boundary conditions become amplified in effect over time. In general, complex or "self-organizing" systems display second-order emergence.

- Third-Order Emergence: Properties emerge as a consequence of shape, time, and "remembering how to do it." Example: Biology, where genetic and epigenetic instructions place constraints on second-order systems and thereby specify particular outcomes called biological traits. These traits then become substrates for natural selection by virtue of the fact that 1) their instructions are encoded and 2) they endow organisms with adaptive properties.

Biological Traits as Emergent Phenomena

It is important at the outset to expand upon this concise summary of third-order emergence and, in particular, set forward what is meant by trait, natural selection, encoding, and adaptation.

Biological traits are made up of biomolecules, like enzymes and hormones and ion channels, that interact and play out in space and time. The difference between traits and complex systems is that the traits are specified by instructions. The shape of an enzyme, and its capacity for productive shape changes, and the timing of its appearance in a given cell, and how much of it is made, and what regulates its interactional possibilities — these things are not left to chance or to fluctuating initial conditions or boundary conditions. They are encoded, either in the genomic (DNA) instructions themselves or in epigenetic instructions (cell-cell interactions), such that pretty much the same outcome — the same emergent trait — occurs with a quite remarkable degree of reliability. And indeed, to generate a reliable outcome is what organisms are about. When a species is unable to reproduce itself in a reliable fashion, it either drifts towards extinction or, via mutation and natural selection, evolves a more reliable strategy.

Granted that the ultimate substrate for natural selection is the organism itself, the units of selection are biological traits. Thus natural selection does not "see" the enzymes, the individual gene products, that catalyze an organism's energy transduction. Rather, natural selection "sees" the outcome, the emergent trait we call metabolism. In the same way, natural selection "sees" an organism's motility and not the contractile and regulatory proteins that together allow that motility to happen. Instructions for a less adaptive metabolism or motility are less likely to spread through a population than instructions for a more adaptive metabolism or motility, with the wild-card word "adaptive" having everything to do with the match between an organism's genomic expectations and the niche wherein it in fact finds itself. Metabolism and motility are nothing but their constituent parts. But they are also something more, something new and emergent. And they are the "stuff" of what an organism is.

With this much background, we can re-visit the definition of third-order emergence using motility as an example. Our muscles contract and relax by virtue of regulated interactions between two kinds of proteins, actin and myosin. Muscle motility can therefore be said to be *nothing but* actin/myosin interactions that generate the ability to move (*something more*). The actin/myosin system self-organizes in muscle cells using second-order "rules" which, in turn, are constrained by the first-order shapes of the participating proteins. That there exists a selectable biological trait called actin/myosin-based motility is dependent on, and the consequence of, there being third-order genetic instructions that specify and constrain and transmit the parameters for such self-organization.

Motility *per se* is not an exclusive property of actin/myosin systems. A number of self-organizing systems in the biological world — the bacterial flagellum, the eukaryotic cilium — generate motility using proteins and mechanisms that are very different from actin/myosin. Thus motility is independent of the *nothing-buts* that serve to generate it, which is true of emergent properties in general (surface tension is a property not only of liquid water but of liquids in general). Motility emerges again and again during evolution because its acquisition is often adaptive for organisms and hence it is subject to positive natural selection.

The larger point, then, is that third-order systems, by being remembered/selected and not simply the episodic outcome of unspecified initial and boundary conditions, have the all-important property that they are subject to constructive influence. We can talk about evolutionary "improvements" in motility — about how a particular kind of motility became better adapted to particular niches via mutation and natural selection of participating proteins — in a way that we cannot talk about improving surface tension or improving the meteorological ramifications of a butterfly flapping its wings in Japan.

Importantly, then, the onset of third-order emergence defines the onset of telos on this planet and, for all we now know, in the universe. Creatures have a purpose, and their traits are for that purpose. What's particularly important about biological traits is that they are about something. Metabolism allows an organism to carry out its chemistry; motility allows it to move towards food and mates and away from toxicity and predators. There is a point to a trait that we cannot ascribe to a snowflake. A trait, and the collection of traits that it combines with to generate an organism, has a purpose, namely, to allow the organism to carry on and thereby transmit the instructions. Organisms of different sorts may inhabit other planets in the universe, but the organisms on this planet, and their inevitable evolution given its inhomogeneous environment, are steeped in teleology.

Emergent Semiotic Systems

Biology is not only a physical/chemical science but also a semiotic science, a science wherein representation and significance are central elements. Semiotic systems, by definition, are emergent: virtually any material property (*nothing but*) can become endowed with semiotic information (*something more*). Given our goal of understanding something about the nature of human minds, it is apt to begin by considering how semiotic systems undergird the construction and perpetuation of biological organisms in general.

At the heart of any semiotic system is the "sign" relationship — the ability of something to "stand for" something else, to "mean" something else, to carry *interpretable* information. So, in DNA, the codon ATG *means that* the amino acid methionine should be placed in a particular position in a protein. The hormone insulin, binding to and activating a receptor on a fat cell, *means that* blood sugar levels are high. And a molecule diffusing from a decaying food source and binding to and activating a receptor on the surface of an amoeba *means that* the food source is nearby. The molecule is not the food source itself but rather a sign indicating its proximity. In each case, a sophisticated biochemistry is recruited to translate/interpret the sign's meaning: Numerous "translation factors" (ribosomes, transfer RNAs) are involved in going from ATG to methionine, and complex signal-transduction pathways instruct the fat cell and the amoeba that certain information has been perceived and that certain responses are indicated.

Awareness in Single-Celled and Multi-Celled Organisms

At this point we need to lift up some important distinctions between the awareness systems of single-celled vs. multi-celled organisms.

Most organisms on the planet today, and doubtless the only organisms on the planet for the first several billion years, were single-celled organisms, of which bacteria, yeasts, and amoebae are the most familiar. Each inherits a genome (a collection of genes) that specifies traits suited to negotiating the niche that the organism expects to encounter, and if the match between genome and niche is in fact is a good one, the organism will be able to grow, copy its genome, and divide into two daughter organisms with one genome apiece. Many of the encoded traits make use of receptors (like the amoeba's receptors for decaying-food molecules) that detect relevant signs in the environment and convey their meaning to the organism. These systems can be said to mediate *cellular awareness*.

Multicellular organisms, originating at least 600 million years ago, partition out the job of being alive to two different kinds of cells: The germ-line cells (eggs and sperm) engage in transmitting genomes to daughter generations, and the remaining somatic cells engage in growth and niche-negotiation. The somatic cells, in turn, go on to sub-specialize in the execution of particular traits — fat cells specialize in glucose storage and heat insulation, muscle cells specialize in motility, and so on — and each is again studded with receptors — insulin receptors on fat cells, neurotransmitter receptors on muscle cells — that mediate cell-type-specific modalities of cellular awareness.

The amoeba, then, is basically a one-man band, whereas a multicellular organism is a very large orchestra. Orchestras require conductors, and while the task of coordinating the traits of a multicellular animal is carried out at many levels, the brain is unquestionably the maestro. The brain receives a vast array of inputs/signs about the environment via various kinds of sensory neurons; it also receives a vast array of inputs from the rest of the body about "how things are going," signs — often hormones or neurotransmitters — that *mean* pain or hunger or fear or sexual attraction; it then integrates this information and oversees the resultant responses we call emotionally-valenced behavior. We can call these semiotic feats *brain-based awareness*.

Particularly distinctive about brain-based awareness is its *indexical* semiotic capacity. When a sensory system is stimulated, it proceeds to make synaptic connections in the brain with 1) neural pathways that encode memories of previous encounters with that kind of stimulus, 2) pathways that encode its various emotional and instinctual valences, as well as 3) numerous learned associations between that stimulus and others that impact on its meaning. So, for a dog, the visual stimulus of a food dish will elicit all manner of memories (previous meals), instincts (hunger), emotional states (anxiety about being hungry, anticipation of pleasure in being fed) and learned associations (the human food-provider, the sound of the food bag opening) that are brain-integrated and then converted into some sort of coherent behavioral response.

In the end, brain-based awareness is *nothing but* cellular awareness. Each neuron is a cell, and neurons utilize the same kinds of receptors and hormones and signal transduction pathways to mediate perception and synaptic transmission that are found in single-celled organisms and in fat cells and muscle cells. But brain-based awareness is quintessentially also *something more*: The indexical possibilities of learning and memory and emotional valence are in theory limited only by the kinds of stimuli that a brain is equipped and motivated to perceive. With all due respect for the highly adaptive hormonal strategies used by plants to integrate their

multicellularity, we would say that it is not just anthropocentrism that motivates our admiration for brains. Brains are inherently amazing.

How Do Mammalian Brains Change?

Not yet mentioned, but key to our story, is the fact that during the course of evolution, brains underwent a major transition in their mode of genetic specification. Whereas the brain of each and every worm of the species *Caenorhabditis elegans* contains 302 neurons that are found in identical locations and mediate identical synaptic pathways, the brain of a mammal contains some 100 billion neurons, plus or minus, whose locations and synaptic relationships are established "on the fly." Genetic scripts endow the neurons that grow up into the embryonic cranium with general instructions as to where they are going and what kind of neurotransmitters they are able to produce, but most of what happens after that is elicited by the other neurons they encounter, the growth factors they secrete and perceive, and the signals they transmit to one another as they make contact and move past. When a neuron picks up on a developmental cue and differentiates along the lines indicated by that cue, its resultant properties then influence the neurons with which it next interacts. That is to say, mammalian brain development is robustly "epigenetic": genes set the process up and continuously participate in differentiation events, but most of the information is exchanged at the level of cell-cell interactions.

The epigenetic course of brain development is clearly reliable: if one were to examine 100 fetal mouse brains, one would find their overall organization to be strikingly similar. But it can also be said to be underdetermined in the sense that if one were able to analyze any two of these brains at a neuron-by-neuron level, there'd be lots of differences —as contrasted with the brains of two worms.

When a developmental process is as underdetermined and epigenetically encoded as is mammalian neurogenesis, then small changes can generate major differences in outcome. That is, the process takes on features of second-order emergence — what happens next can be highly influenced by what has happened before — and simple mutations can have large-scale downstream consequences and hence large-scale evolutionary consequences.

We can consider three ways that this can happen.

- Parts of the brain may *change in size*. A single gene mutation can result in an increase or a decrease in the number of cell divisions that a given lineage undergoes during the course of embryology. Mutations of this kind routinely generate heritable differences in the overall size of an organism, and the evolutionary consequences of such changes can be significant. This is particularly true for loosely-determined mammalian brains, where an additional doubling of certain neurons provisions the brain with a "new set of players," and hence new connective opportunities, while a halving of such neurons means that their former potential synaptic partners will probe the "brain space" in search of new connective opportunities.

- Brain pathways may *degrade* when they are no longer under selection. For example, when fish or mammals come to inhabit caves or underground niches where there is no light, natural selection no longer operates to maintain their visual systems. Therefore, mutations that compromise visual acuity are not selected against — no one can see anything anyhow — and

the animals eventually become heritably blind.

- Degraded pathways may be *selected to reconfigure*. Figure 1 compares the brain of a sighted rodent and a blind mole rat. In the sighted species, the visual cortex, heavily innervated with optic-nerve input, mediates vision, whereas in the blind species, lacking optic input, this same region of the brain has been "taken over" by neurons delivering auditory and tactile input. Not only does this illustrate the underdetermination of mammalian brains — there is no hard-wired "visual cortex" *per se* but rather a cortical region that is induced to mediate vision when programmed by optic input and to mediate touch and hearing when programmed by tactile and auditory input. It also illustrates the strong role that natural selection can play in shaping brains, since the enhanced hearing and touching afforded by these new cortical connections are presumably adaptive for the blind mole rat in negotiating its underground niche.

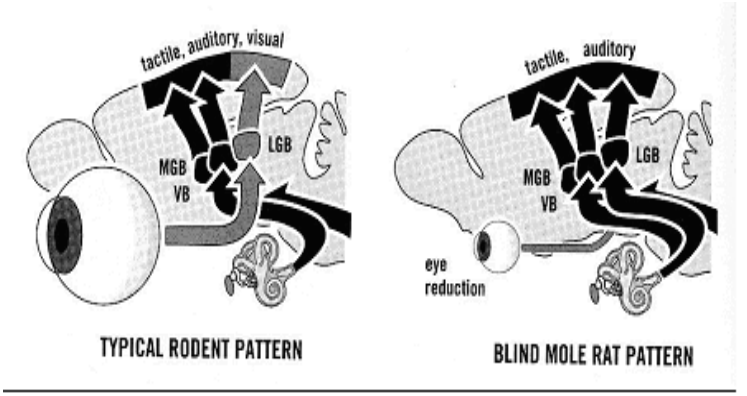

Figure 1 Innervation of the cortex in a typical rodent and in a blind mole rat inhabiting an underground niche. (From Deacon, T., *The Symbolic Species*, 1997*)*

Mental Traits Shared by Human and Non-Human Primates

So what about human brains?

Figure 2 illustrates our family tree. Some 5 million years ago — a short span in the ~600 million years of animal evolution — a common ancestor gave rise to 3 lineages: the human, the chimpanzee, and the bonobo. The mental traits shared by these 3 kinds of animals can be assumed, as a first pass, to have been present as well in our common ancestor, whereas traits found in only one lineage can be assumed, again as a first pass, to have evolved as lineage-specific events.

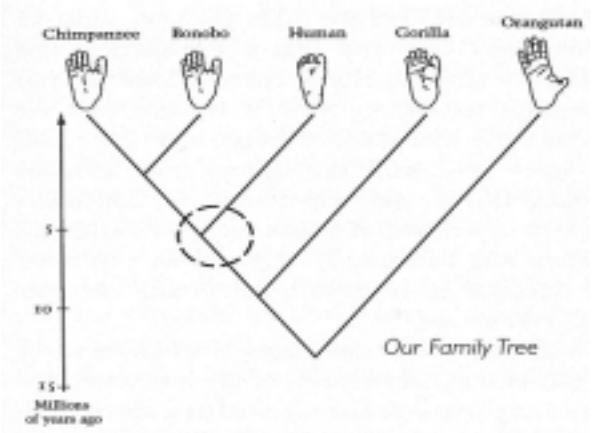

Figure 2. Family tree of the great apes Common ancestor is circled. (From Wrangham, R. and D. Peterson, *Demonic Males*, Houghton Mifflin, 1996)

Humans, chimps, and bonobos share numerous mental traits which, by this reasoning, would have been displayed as well by our common ancestor. We are all highly intelligent animals with impressive abilities to learn by both experience and imitation and to remember what we have learned. We display a similar range of temperaments generated by similar emotional systems: primatologists who come to know chimps and bonobos can readily describe one as shy, another as extroverted, another as impatient, and so on. We are all similarly dependent on maternal care and nurture for appropriate mental development, with deprivation generating a similar syndrome of impairments. And finally, we are all social animals, living in highly structured groups with similar organizing principles. As considered more extensively in ref. 3, these include a robust attention to *social hierarchy*; a preoccupation with the *nurture* of the young (also called kin altruism); skillful engagement in *strategic reciprocity* ("I'll scratch your back if you scratch mine"; also called reciprocal altruism) and the attendant formation of friendships and alliances; a hostility towards outgroups (*xenophobia*) and an endowment of the pro-social capacity we can generically call *empathy*.

Preston and de Waal have recently published a comprehensive review of the evidence for empathy in non-human primates (4), and the findings are convincing: Co-existent with the self-interest inherent in all organisms and necessary for their survival (3), and co-existent as well with such "negative" capacities as aggression (considered briefly at the end of this article), primates are disposed to help one another out in the service of group stability, to be tolerant, to offer forgiveness and consolation and forge reconciliation,. Countering arguments that "being nice" fails to foster transmission of one's genes and that any genetic disposition towards empathy would be quickly trumped by winner-take-all "cheats" is the compelling argument that when survival is dependent on group coherence, as is by definition the case for robustly social animals like the primates, there occurs positive selection for the capacity to sense and respond to the emotional status of others in the troop, and negative selection against sociopathic behavior.

Mental Traits That Are Uniquely Human: A Scenario for their Emergence

It goes without saying that humans have many unique mental traits, but we would argue that one trait is foundational to the rest: Humans learn not only by imitation and experience but also by accessing information from cultures that are encoded in symbolic languages. It's not so much that we have more of chimp-like intelligence; rather, we also have a different kind of intelligence.

In the following paragraphs we present a scenario for the co-evolution of language, culture, and symbolic human minds; many of these ideas are presented in greater depth in refs. 1 and 5. The scenario is by definition a speculation — what actually happened may never be fully known — but we find the scenario heuristic, helping us to focus in on what's distinctive about human mentality, and some of its propositions should eventually be amenable to empirical evaluation.

Niche Construction

To understand the role of culture in human evolution, it is helpful to start with the beaver. Beavers exhibit the remarkable trait of damming up streams to form ponds and then inhabiting the ponds, thereby protecting themselves from predators. This trait is clearly "hard-wired:" If the sound of running water is broadcast to captive beavers, they proceed to pile sticks on top of the speaker. It also exemplifies *niche construction*: beavers are adapted to the ponds that they themselves create; they are selected for their ability to produce the niches upon which they depend.

And so it is with the human. The niche to which we are adapted — human culture — is a niche that we ourselves construct; we are selected for our ability both to produce and to inhabit culture-based niches. Since human culture is encoded in and acquired by symbolic languages, this means we have been selected for our symbolic minds in the same way that beavers have been selected for their dam-building skills.

Co-Evolution of Culture, Language, and Brain

Culture, language, and symbolic brains that manipulate language have co-evolved, by our scenario, in a constructive, accretive fashion: If a more facile symbolic manipulation were made possible by a new kind of brain configuration which in turn allowed better access to adaptive kinds of culture-based understandings, there would occur selection for such a trait which, in turn, would generate hominids yet more dependent on culture for survival and hence more likely to be selected for further "improvements" in their language facility. While we have no speculations to offer here as to why/how this might all have gotten started in the first place, once a co-evolutionary cycle like this gets set up, it can take on a life of its own and can evolve very rapidly.

This being said, there remain many questions as to how it all might have played out. We next suggest a dynamic that could have been operant.

Masking, Degradation, and Reconfiguration

As we saw in our story of the blind mole rat, a trait that is no longer under selection, like vision in darkness, is prone to degradation. This is not because the darkness "causes" the degradation, of course. Mutations leading to a loss of function occur on a regular basis — most changes in information systems result in information loss, as *per* the second law of thermodynamics. As long as the trait is under selection, degradative changes are winnowed out, but when selection backs off, they persist and accumulate.

Selection can become attenuated when the trait is no longer needed, like vision in darkness. It can also attenuate when the trait's function comes to be provided by the environment, rendering the trait redundant, a dynamic known as *masking*.

To illustrate masking, we can consider the story of how ascorbic acid came to become vitamin C. Most organisms are genetically programmed to synthesize their own ascorbic acid, which is necessary for their survival. Ancestral apes in our lineage, however, started to eat ripe fruits, which are rich in ascorbic acid. As a consequence, mutations compromising the enzymes involved in ascorbic acid biosynthesis were not debilitating because abundant ascorbic acid was already coming in from the diet; that is, the mutations were masked from natural selection. Hence the pathways degraded, and ascorbic acid became a vitamin: Chimps and humans now *must* obtain it from the outside.

Applying this concept to human evolution, we can posit that to the extent that culture came to provide hominids with useful information from the outside, any genetically established programs specifying overlapping kinds of information in the brain would be similarly masked from selection and would therefore become prone to degradation. As this occurred, hominids would become increasingly dependent on — indeed, addicted to — cultural information for their survival.

We can now circle back to our consideration of brain evolution, recall that degraded brain programs tend to become reconfigured to support alternative adaptive traits, and posit that parts of the hominid brain became reconfigured for language. Since the neural basis for language capability is not now understood, it is not yet possible to point to novel features of human brains, and absent from chimp brains, that illustrate the physical basis of this emergent skill. Still, it is by definition the case that to the extent that any degraded programs became re-configured for linguistic operations, this would allow better access to the cultures upon which hominids had become dependent, meaning that such degradation/reconfiguration events would be adaptive and subject to positive selection.

To summarize, then, we can offer an evolutionary model for the emergence of human minds:

- Culture has masked the need for certain genetically-encoded ("phylotypic") primate mental pathways, and these have degraded.

- The freed-up "brain space" has been reconfigured to generate minds adept at learning symbolic language and hence acquiring cultural information.

What Does Symbolic Language Accomplish?

So what's so special about symbolic language? One of us has written a 500-page book (1) on this subject; therefore, in lifting up a few key concepts , we are by definition leaving out most of the story.

Symbolic representation is a novel emergent semiotic capacity found only in the human (and in machines designed by humans). The *nothing but* is the indexical primate brain, highly sophisticated and versatile, that we admired earlier. The *something more* is the ability of the human brain to use symbols (*words*) to refer to indexes and to sets of indexes, and to use *syntax* to indicate the relationships of these words to one another. As stressed in *The Symbolic Species*, these operations are not likely to have evolved, as some have proposed, *via* adding some "language box" to a primate mind so that it can manipulate languages that are somehow "out there" to be discovered. Rather, there likely occurred a co-evolution of language and brain such that symbolic languages were selected for their ability to be learnable by children's brains and children's brains were selected for their ability to learn symbolic languages. If we add to the mix the proposed selective effects

exerted by niche construction and cultural addiction, and the proposed degradation/reconfiguration dynamics facilitated by cultural masking, we come away with the conviction that the evolution of human minds can be modeled in plausible evolutionary terms, and hence need not be relegated to the miraculous or mysterious.

Nevertheless, if brains are amazing, the human brain is flat-out astonishing, most especially because the language user can generate constructions that are independent of "lower" levels of reference, meaning that *the words themselves come to define a virtual reality that has a life of its own.* When challenged we can pause, "unpack" our symbols and syntax, and examine their antecedents, but for most purposes we inhabit the linguistic constructions themselves.

Words can function as straghtforward signs, but most words can also be used to encode and convey what we call concepts or ideas, pointing to complex sets of indexes that are integrated via complex syntactical relationships. Moreover, since these concepts inhabit a virtual reality with a life of its own, we can and do engage in *conceptual blending*, mixing and matching and transfiguring concepts in an orgy of semiotic freedom, constrained only by whether what emerges has some sort of meaning (the definitional constraint on semiosis). As our cultures have evolved to harbor and convey increasingly sophisticated concepts and ideas, moreover, the criterion as to what-has-meaning has also lost its obligation to real-world antecedents: Whole new kinds of meaning permeate the virtual world that we call the imagination.

And now, a central claim. We would argue, as have many others, that our sense of "self," what we will call human self-awareness, is made possible by symbolic language. That is, when we say that we are aware of our thoughts and ideas and plans and memories, we do this using symbolic constructions. It may be possible to have a thought without linguistic representation, but we only know that we have had one when it is self-represented in symbolic form. This claim is made in full awareness of its attendant ambiguities, such as "how do we know that a dog is not also self-aware?" or "what about the pre-linguistic human infant?" While we find these questions intriguing, they fail to vitiate our sense that once language is in place, there emerges not only symbolic reference but also symbolic self-reference in the sense that we humans experience that experience.

How symbolic self-reference "works" is as elusive as how language itself "works." But if neuroscientists were tomorrow to publish a definitive description of the biophysical/neural basis for human self-awareness, the account would be unlikely to have much impact on our understanding of our mental theaters because we are already expert at what they are like. Moreover, our self-awareness *seems* to be independent of mechanism. As much as we are able to acknowledge the embodiment of our ideas and feelings, we experience them as operating in a disembodied virtual realm.

Emergent phenomena are not prefigured. They come for free, apparently out of thin air. And so it is with our selves. Our thoughts and actions are determinate but not predetermined. Self and experience are both entirely physical and entirely representational. What a person is and what a person is conscious of are representations, and representations — although *nothing but* physical objects and events — are *something more* as well.

Our very selves, then, are by this rubric emergent phenomena. It is the very possibility of being a locus of experience, and feeling, and perpetual coming into being, that is a person. This emergentist view of who we are, neither radically reductionist nor setting us apart in some disconnected realm, is for us a thrilling way to ground our existence.

Experiencing Our Primate Minds Symbolically

No question about it: Our symbolic minds allow us to access mental experiences, like mathematics and aesthetics and spiritual intimations, that we have every reason to believe are novel to the human, unique to the human. Our poets, artists, philosophers and religious leaders provision us with rich and provocative descriptions of these experiences, and our cultures allow us to transmit, retrieve, and build upon their seminal insights. In what follows we are in no way suggesting that these insights are not of utmost importance to what it means to be human.

But we suggest that it is also of utmost importance that we not lose track of our mental evolutionary antecedents. To say that our brains have undergone critical reconfigurations as they evolved their capabilities for symbolic (self)-representation is not to say that our common-ancestor brains were left in the dustbin. As noted earlier, we share strong cognitive and emotional homologies with our primate cousins, and to the extent that degradation/reconfiguration went into generating our capacity for language, it occurred in the midst of a primate brain that remains very much a primate brain. Any perspective on the human condition that brushes this fact aside is an incomplete perspective — indeed, we would say that it is an impoverished perspective.

A common response to this interface is to propose a *de facto* dualism. Yes, it is acknowledged, much of who we are has primate antecedents, but, given our emergent minds, our rationality, our spiritual yearnings, and our culturally-encoded meaning systems, we somehow have the wherewithal to transcend these antecedents and operate in a set-apart matrix of human-specific truths. Indeed, this dualism is inherent in the claim, loosely called the naturalistic fallacy, that *you can't get an ought from an is*. We may well inherit an (often awkward) evolutionary legacy, but it has no *a priori* claim on our modes of valuation and, in particular, on our ethical codifications.

An alternative to such forms of dualism, and one that we find more germinative and satisfying, is the notion that one of the things that we do with our symbolic minds is to *experience our primate minds symbolically*. Our primate minds have not gone away (albeit some phylotypic "instincts" have been lost and perhaps reconfigured), nor are they experienced as apes would experience them. They are experienced as experienced by human minds: symbolically.

This notion can be fruitfully applied to many traits, e.g. our experience of sexuality. In the remaining sections of this paper we will develop the notion in the context of morality. The thesis: Given that we have evolved from an intensely social lineage, we are uniquely aware of what it feels like to be pro-social, and it is this awareness of what it feels like to be moral — this moral experience — that undergirds and motivates the actions of a moral person.

Moral Experience

Moral experience, we suggest, entails a coupling of our rich heritage of social orientation with our ability to symbolically represent it to ourselves. During this coupling, the experience of our pro-social capacities, and their role in affecting action, is radically transformed, and what emerges is a major augmentation of our social heritage. We are able to apply these amplified pro-social capacities to experiences and imaginings and modes of action that are no longer constrained by evolutionary precedents and classes of phylotypic stimuli. Indeed, our capacity for conceptual blending allows a synthesis of moral understandings and emotional experiences that

would otherwise be mutually exclusive.

It follows that morality is not something that humans acquire via cultural instruction, albeit, as we discuss later, culture serves to complement the process in important ways. Rather, we are led to moral experience and insight. Real morality can't be forced on people, nor can they be fooled into having it, nor do they just act on their 'moral instincts.' Real morality does not simply bubble up from beneath, nor is it imposed from the outside. In each one of us, it must be discovered anew. The discovery process may require great mental and emotional effort and may bloom only in the right climate, but human beings see morality, recognize it, regardless of what it is that they want or need or love or hate or feel compelled to do.

We can put flesh on the abstractions by considering the psychopath. A psychopath can negotiate hierarchy and execute strategic reciprocity without difficulty, and can learn, and simulate, moral behavior when this suits his purposes. But, be it by inborn error or brain injury or childhood deprivation, he lacks the capacity to experience moral experience, to feel anything in the way of empathy, to put himself in another's shoes. Morality without empathy is by definition oxymoronic. Therefore, his simulation of morality is strictly instrumental, and, in extreme cases, he is able to say things like "I killed that kid because I'd never killed a kid before and I wanted to see what it felt like." The tragedy of the psychopath reminds us that without access to moral experience we are no longer fully human.

Virtues, Pro-Social Orientation, and Moral Experience

The notion that human morality is located within moral experience is not a new insight. It is embedded, for example, in the thinking of Aristotle, who wrote: "We have the virtues neither by nor contrary to our nature. We are fitted by our nature to receive them." Subsequent philosophers have continued to explore this approach, developing a tradition known as virtue ethics (see, e.g., ref. 6).

So what are the virtues, and how do they relate to the thesis that morality entails the human experience of pro-social orientations? Four of the virtues that appear on most lists — humaneness/compassion, fair mindedness, care, and reverence — can be thought of as related to four of the inherited pro-social capacities that we listed earlier — empathy, strategic reciprocity, nurture, and hierarchy (see also refs.3 and 7). We develop these correspondences briefly below in order to indicate how this line of thinking might be pursued.

Morality without empathy is oxymoronic, as we have said, and the words humaneness and compassion are among those used to describe the emergent way that humans access, experience, and manifest the empathic nature inherent in our heritage. We come to grasp that to put oneself in another's shoes is not only something that applies to our kin or friends or social group. Indeed, as our vocabularies mature and our ability to manipulate concepts complexifies, we become able to articulate empathic connection with such abstractions as "life itself" or "the planet Earth." Moreover, we can engage in conceptual blending and configure empathy in radically new ways, as in "Love Thine Enemy."

If humaneness/compassion can be said to entail the symbolic accession of empathy, then fair mindedness strikes us as entailing a symbolic synthesis of humaneness and strategic reciprocity. Once someone takes on board the notion that for every winner there is a loser, and once someone has the experience of putting oneself in the shoes of the loser and caring about her, it becomes evident that there is another way to think about these social interactions, namely, that more important than winning or losing is that the outcome be fair. Strategic

reciprocity fused with humaneness emerges as a sense of justice, a centerpiece of moral philosophy. And again, we can complexify further and articulate a sense of ecological justice.

The third virtue, care, inherent as well in such concepts as responsibility, commitment and kindness, emerges from the strong primate sense of nurture, not only of one's own offspring but also, in a lineage wherein paternity is uncertain, of all the youngsters in the troop. Primate nurturance entails not only protection and provisioning but also relationship, play, and affection, and this capacity, we suggest, transfigures as the capacity to care about one another and about larger concepts like ecosystems and future generations. Compassion and care overlap, but care is the more active noun and emerges, we believe, from distinct primate antecedents.

The fourth virtue, reverence, can be modeled as a complex emergent manifestation of our orientation in hierarchy. Reverence, in its mindful manifestations (7), describes the capacity to carry the sense that we inhabit contexts that are larger and more important than ourselves, to which we accord awe and respect and gratitude. We come to speak of reverence for our leaders, and leaders to speak with reverence of their followers. We orient ourselves in reverent family life and reverent communities, and offer honor to revered understandings in ceremony and ritual (8). Many find orientation in a theistic reverence, while others become besotted with reverence for the natural world, the emergent material world, in all its wondrous manifestations and evolutionary history (9). The human capacity for reverence, we suggest, may represent a transfigured version of our innate grounding in social valuation, endowing us as well with a sense of humility that allows us to ward off the perils of hubris.

Moral Motivation

To have moral experience is, of course, quite a different matter from acting in a moral way, particularly when it is against one's self-interest to do so. We may see what is right but not be motivated to act on it.

The all-too-common practice, now and probably throughout human history, is to provide moral motivation by rewarding "good" thoughts and behavior and punishing "bad," as in "Santa knows if you've been good or bad so be good for goodness sake" or "If you do that you will be punished by the gods/ancestors." This practice turns morality into a commodity that can be bartered, a substrate for self-interested strategic reciprocity, an entity that fills the Christmas stocking or assures a glorious afterlife. The problem with this, of course, is that humans quickly notice that there are other strategies that also fill the stocking, like deception and greed, and that these are in fact more reliable strategies. The commodification of morality is, to our mind, one of the most dangerous things that we do, quite as dangerous as fundamentalism or moral relativism.

But if moral motivation is not to be provided by punishment/reward systems, then where is it to come from? Aristotle makes an interesting claim here, which is that "virtuous conduct gives gladness to the lover of virtue." Note that he is not saying that virtue brings gladness to the virtuous, but rather to the lover of virtue.

One way to think about Aristotle's claim is in the context of what 17th century philosophers like Shaftesbury and Hume called moral beauty. The idea is that we access and enjoy moral beauty along the lines that we access and enjoy aesthetic beauty, where in both cases the rewards are both private and ineffable. Importantly, the lover of virtue is made glad not only by experiencing moral beauty in himself, which could carry a lurking reward motivation as in "this will assure my place in heaven," but also by witnessing moral beauty in others – New York firefighters, for example, or persons who reach a fair outcome to a conflict. We say that moral

experience "warms the heart," often reflexively placing our hand over our heart as we say it; we say that we are uplifted. Indeed, those who self-identify as worldly sophisticates may feel somewhat sheepish to find their eyes filling with tears at some experience of moral beauty, and this can be dismissed all too quickly as sentimentality. Before dismissing sentimentality, we might first want to deepen our understanding of what it entails.

To invoke as a moral motivator the heart-warming sense of gladness that we experience when we encounter moral beauty is, on the one hand, to say very little since we know so little about what it means to perceive beauty, be it aesthetic or moral. But we do know that we seek such experiences and find them meaningful, and to our mind there is much to explore along these lines, particularly from the perspective of helping our children to access morality for its attendant sense of beauty rather than because it promises a full stocking.

Moral Ideals

Our focus on the "bottom-up" sources of moral experience has seemingly ignored our earlier focus on the importance of human culture, a deficit we will now address.

Cultural traditions include the writings of numerous philosophers and theologians who derive moral constructs from a priori rational premises and offer resultant codes of ethical conduct. Many of their insights and codifications — the Golden Rule, the Categorical Imperative, the Veil of Ignorance, the Eight-Fold Path — robustly complement the understandings that are accessed during the process of moral self-discovery. But we would suggest that the core contribution of culture along this axis is that it encodes and presents to us moral ideals that guide our moral maturation and stimulate our moral motivation.

Moral ideals come to us in artistic/narrative form. We hear stories or see paintings or sing songs about people who are good, who do the right thing at the right time in the right way, and we lock in, we sense the correspondences with our own pro-social biases. We are "inspired" to be like them.

All religious traditions, throughout the ages, rely on artistic narrative to convey moral ideals, to educate the emotions. Moreover, these narratives function independently of the metaphysical claims of the tradition: a Christian has no problem accessing the compassion that inheres in the images and stories about the Buddha, nor the reverence that permeates a Native American tribal ceremony. Indeed, a recent survey of world religions reports a deep congruence in moral ideals despite vast differences in metaphysical premises (10). And while, of course, religious institutions, like all institutions, are vulnerable to being hijacked under stressful circumstances into advocating the likes of violence and cruelty, they return to their pro-social narratives once the stressful circumstances abate.

Moral experience, we suggest, is the wellspring of our virtue — without it we are doomed to psychopathology — but once it is perceived, which seems to begin early in childhood, we embark on a lifelong journey, fraught with encounters with fear, greed, hubris, prejudice, and self-absorption, wherein we seek to act in accordance with the beauty of the good. This journey is described, in countless metaphors, by our religious traditions, and whether persons encounter these metaphors as the word of the gods or, as we do, the word of the best that resides within the human, our journey would be barren without them.

Most importantly, moral experience is not only something that we develop within our own beings. We also share this experience with one another, and it binds us together. There are many ways that communities are held together via "straight" kin altruism and hierarchy and strategic reciprocity; indeed, these are robustly operant in our political and economic forms of social stabilization. But our shared moral experiences generate as well a thirst for moral communities. Humaneness, fair-mindedness, care, and reverence can be considered to represent cardinal virtues in the sense that a human community mindfully infused with these qualities can be described as a moral community — within which, we believe, can best flourish our emergent, and most astonishing, minds and selves.

A-Sociality

To look at the primates and lift up only their pro-social capacities is, of course, to tell only part of the story. Always central to our evolutionary nature is our self-interest, and always lurking in the wings of self-interest are its "darker" manifestations. It is here that the project of naturalizing morality encounters for many its insurmountable hurdle. When we remember that apes are also observed to injure and even kill one another, to use force in sex, to be cruel and rejecting, and to display robust xenophobia, we become distinctly uncomfortable, and invoke with a shudder the specter of the criminal basing his legal defense on the claim that "my genes made me do it."

A full consideration of the interplay between self-interest and pro-sociality, particularly as each plays out in its emergent manifestations, is well beyond the scope of this essay, but a few observations are germane. First, it is important to point out that the existence of self-interest, and its darker forms however defined, does not negate the existence of pro-sociality. Pro-social capacities are not just the absense of a-social capacities. They have lives of their own.

We can then recall that primates, both nonhuman and human, most often engage in a-social behaviors when they are subjected to stress, and particularly to prolonged stress. Under these circumstances, we hunker down and engage in self-interested survival patterns, the default behavior of all creatures, and these often take forms that are antithetical to pro-sociality. One way to stack the deck in favor of morality, therefore, is to ameliorate the conditions wherein humans find themselves physically or emotionally impoverished, threatened, defeated, abused, humiliated, lonely, or insecure. Such conditions foster the dehumanization and demonization of those identified as the "cause" of our frustrations, allowing them to become targets of exclusion and brutality (11). Such conditions also render humans vulnerable to rigid fundamentalisms — many carrying a morality label — that activate our fear and greed in their promises of deliverance.

References

*A version of this essay was presented at the November 2002 annual meeting of the Polanyi Society. This version of the essay was originally published in *Zygon-Journal of Religion and Science* 38:4 (Dec. 2003): 801-819 and is reprinted with permission.

1. Deacon, T.W. 1997. *The Symbolic Species: The Co-Evolution of Language and the Brain.* New York: W.W. Norton.

2. Deacon, T.W. 2003. "The hierarchical logic of emergence: Untangling the interdependence of evolution and self-organization." In: B. Weber and D. Depew, eds. *Evolution and Learning: The Baldwin Effect Reconsidered*.

Cambridge: MIT Press.

3. Goodenough, U. 2003. "Religious naturalism and naturalizing morality." *Zygon*, in press (March edition).

4. Preston, S.D. and F.B.M. de Waal. 2002. "Empathy: Its ultimate and proximate bases." *Behavioral and Brain Sciences*, in press. Online: www.bbsonline.org/preprints/preston

5. Deacon, T. W. 2003. "Multilevel selection in a complex adaptive system: The problem of language origins." In: B. Weber and D. Depew, eds. *Evolution and Learning: The Baldwin Effect Reconsidered*. Cambridge: MIT Press.

6. Hursthouse, R. 1999. *On Virtue Ethics*. New York: Oxford University Press.

7. Goodenough, U. and P. Woodruff. 2001. "Mindful virtue, mindful reverence." *Zygon* 36:585-595.

8. Woodruff, P. 2001. *Reverence: Renewing a Forgotten Virtue*. New York: Oxford University Press.

9. Goodenough, U. 1998. *The Sacred Depths of Nature*. New York: Oxford University Press.

10. Ferrer, J.N. 2001. *Revisioning Transpersonal Theory: A Participatory Vision of Human Spirituality*. New York: State University of New York Press.

11. Glover, J. 1999. *Humanity: A Moral History of the Twentieth Century*. New Haven CT, Yale University Press.

[**Ursula Goodenough** (ursula@biology2.wustl.edu) is Professor of Biology at Washington University, St. Louis. She has served as president of the American Society of Cell Biology as well as the Institute on Religion in an Age of Science (IRAS). She is author of *The Sacred Depths of Nature* and many articles on science and religion. **Terrence W. Deacon** is a professor in the Department of Anthropology and the Wills Neuroscience Institute at the University of California, Berkeley. He is the author of *The Symbolic Species: The Co-evolution of Language and the Brain*.]

Polanyi Society Membership

Tradition and Discovery is distributed to members of the Polanyi Society. This periodical supercedes a newsletter and earlier mini-journal published (with some gaps) by the Polanyi Society since the mid seventies. The Polanyi Society has members in thirteen different countries though most live in North America and the United Kingdom. The Society includes those formerly affiliated with the Polanyi group centered in the United Kingdom which published *Convivium: The United Kingdom Review of Post-critical Thought*. There are normally three issues of *TAD* each year.

Annual membership in the Polanyi Society is $25 ($10 for students) beginning in the fall of 2002. The membership cycle follows the academic year; subscriptions are due September 1 to Phil Mullins, Missouri Western State College, St. Joseph, MO 64507 (fax: 816-271-5680, e-mail: mullins@mwsc.edu). Please make checks payable to the Polanyi Society. Dues can be paid by credit card by providing the card holder's name as it appears on the card, the card number and expiration date. Changes of address and inquiries should be sent to Mullins. New members should provide the following subscription information: complete mailing address, telephone (work and home), e-mail address and/or fax number. Institutional members should identify a department to contact for billing. The Polanyi Society attempts to maintain a data base identifying persons interested in or working with Polanyi's philosophical writing. New members can contribute to this effort by writing a short description of their particular interests in Polanyi's work and any publications and /or theses/ dissertations related to Polanyi's thought. Please provide complete bibliographic information. Those renewing membership are invited to include information on recent work.