Biological Considerations in the Analysis of Morality

Richard D. Alexander

> The theory of natural selection identifies the self-interest of every individual organism as the maximal representation of its own genes in future generations. There is no encouragement for any belief that an organism can be designed for any purpose other than the most effective pursuit of this self-interest.
>
> —Williams 1989:196-97

> ... not surprisingly, philosophers and anthropologists have taken sociobiologists to task for their cheerful optimism that the biological reduction of symbolic culture is just around the corner.
>
> —Murphy 1982:93

I will discuss theories and facts from evolutionary biology with which I believe analyses of human activities and tendencies such as moral behavior and thinking must be compatible, and from which to some degree they must eventually derive. I will identify what I see as the most difficult problems or paradoxes for analyses of morality arising out of biological theories and facts, and discuss some possible solutions.

I begin with the assumption, familiar to those who will read this chapter but accepted by only a minuscule proportion of humans in the world, that all organisms, including humans, are evolved through a process guided principally by natural selection, or differential reproduction of genetic alternatives. Directly or indirectly, so necessarily are all of their features, including the ability and tendency of humans to generate, use, and analyze the concepts of ethical and moral and their opposites ("indirectly," here, means that some attributes or behaviors appear as incidental effects of mechanisms or traits evolved in other selective contexts). The arguments need not be repeated; they have been given elsewhere—amply, repeatedly, across decades, and on a wide scale now in even introductory biology courses (Dawkins [1976] 1989; Alexander 1979). The reasons for widespread failure to accept or understand them, among both academics and non-academics, seem not to have received serious consideration. I will suggest that these reasons derive partly from intuitive feelings arising out of human perceptions of motivations of both self and others and the use of intent; for the evolutionary biologist this problem includes the reasons for incomplete control of expression of the emotions and peculiarities of use of conscious and nonconscious "intent" (Alexander 1989). Perhaps as much as the prior existence of alternative explanations (such as divine

creation), or more mundane difficulties of understanding, such feelings may ultimately be responsible for failure of knowledge and acceptance of evolution to spread, and for hostility toward it. I imply that the indifference and hostility of academics and other educated people is a crucial barrier to widespread acceptance of evolution by natural selection as an explanatory device for human self-understanding. One consequence of this hostility and indifference is that anyone can secure the highest possible degree or position in any human-oriented discipline except medicine without taking so much as a single course in the biological sciences, and in medicine without taking a course in evolutionary biology.

Within the past quarter of a century, Darwin's (1859) argument that natural selection is the principal guiding force of organic evolution has been progressively refined. Thus, since Williams (1966), arguments have been widely accepted that natural selection is more potent at lower levels, so that when there are conflicts at different levels the lower levels tend to "win." Essentially this means that the interests of individual organisms tend to prevail as the driving force in the evolution of traits. In 1964, Hamilton presented the first thorough and reasonable discussion of inclusive fitness— the concept that reproductive success by the individual can be achieved by helping collateral relatives as well as by producing and helping descendant relatives. Then Trivers (1971) developed the argument that, because of the possibility of reciprocity, under certain social circumstances selection can even favor helping nonrelatives. Reciprocity can be either direct or indirect, in the latter case involving benefits, such as increased social opportunities, accruing as a result of reputation. Prominence of indirect reciprocity in human social interactions and the importance of reputation creates opportunities for, at the extreme, positive effects on reproductive success from even indiscriminate beneficence (Alexander 1987).

These are the fundamental contributions to current understanding of the working of evolution as concerns social behavior, especially that of humans. They have enabled us to be confident about issues that had been debated endlessly within biology and used uncritically and vaguely both inside and outside biology. That certain opinions on these various topics (both biological and nonbiological)—and others relevant to the following discussion—happened to be correct during previous periods of history (while others were not) gives us no leave to cite their authors as having either demonstrated or understood the essentials of evolution by natural selection prior to the above contributions, or as having employed such knowledge in reaching their conclusions. The only individual who expressed an understanding of the "levels of selection" argument prior to Williams (1957, 1966) appears to be Fisher (". . . the principle of Natural Selection

501

. . . affords no . . . explanation for any properties of animals or plants which, without being individually advantageous, are supposed to be of service to the species to which they belong"—[1930] 1958:49; see also, heroism, p. 264); and only Darwin (1859), Fisher [1930] (1958), Haldane (1955), and Williams and Williams (1957), seem to have shown rudimentary understanding of the significance of reproduction via collateral relatives prior to Hamilton (1964). None of the extensive discussions of reciprocity in the social science or other literature approaches the clarity and perspicacity of Trivers's analysis, or has provided the means for analyzing reciprocity by placing it into an evolutionary context. We can, therefore, reject widespread implications that the above contributions represented nothing particularly new after all. In fact, the effect since 1964 is necessarily revolutionary. [For a compact and easily understood "short course" in modern evolutionary biology—by which I mean ideas and facts developed since 1964, particularly as they bear on social behavior—I recommend Richard Dawkins's three books, *The Selfish Gene* (original 1976, but read the revised edition 1989), *The Extended Phenotype* (1982), and *The Blind Watchmaker* (1986).]

This beginning from biology poses four principal problems for students of morality. First, the potency of selection at lower levels such as the individual questions the usual view of moral behavior as serving the interests of others, or of the group as a whole, and may explain the failure of the seemingly logical extension of beneficent behavior eventually to serve all people equally. Second, there seems to be widespread difficulty in understanding how evolution can produce via differential reproduction of genes a phenotype whose actions are not wholly determined by antecedent ontogenetic events in such fashion as always to yield reproductive advantage. This difficulty seems to many to remove any possibility of an evolved tendency to be moral as a result of choice. Third, given arguments that individuals have evolved to serve their own (genetic) interests and the questions surrounding genetic and ontogenetic determinacy, the problem of understanding conscious intent or motivation versus actual behavior, and its outcomes regardless of conscious intent, is at the very least daunting. · Finally, and perhaps most paradoxical of all, if motivation and expressions of the emotions have evolved to serve the interests of the individual, it is difficult to understand why individual control over the emotions and conscious understanding of motivation are so obviously incomplete or imperfect, or even deceptive, to the actor him/herself—that is, why conscious effort or intent in social communication is often superseded or contradicted by seemingly involuntary or nonconsciously produced transmissions of knowledge and intent. The intuitive skepticism arising out of these problems seems to cause real difficulty in understanding evolution, as it applies to human morality. As said

earlier I regard this skepticism as largely responsible for evolution not being taken seriously by those who do know about it, and by extension responsible for evolution failing to become widely known and widely applied to understanding human behavior.

Levels of Selection, Heredity, and Ontogeny: The Nature of the Organism

Group selection and ontogeny are two of the topics least well discussed, not only by Dawkins in the above-recommended publications, but by evolutionary biologists in general. These two topics, however, must be related to one another in the effort to understand the most important product of evolution (and, therefore, the most important concept underlying a biological analysis of morality), the individual organism. Our failures here are responsible for our inability to make sense of evolution to those attempting to use it to understand themselves and other humans.

Although Dawkins (following Williams 1966) argues successfully that populations and societies of individual organisms are not group-selected, in the main he fails to discuss living entities that *are* group-selected. As a consequence he does not discuss extensively the results of group selection. The genes in the genomes of asexual organisms (whether primitively or secondarily asexual) are in fact group-selected, and for most practical purposes so are those in sexual organisms (Alexander and Borgia 1978). Dawkins (1989:273), in an argument that the organism itself is not a replicator, acknowledges that "In the case of an asexual stick-insect, the entire genome (the set of all its genes) is a replicator." He does not, however, describe the genes in the genome of an asexual organism as group-selected and discuss the consequences of that fact and how they bear on the hierarchical organization of life and our understanding of the organism.

The evidence that genes in genomes are group-selected is, first, that they have evolved to co-exist in the same group (genome), and to die within the group in which they initially placed themselves, leaving only their dispersed progeny in sexually produced gametes or their descendants as products of asexual mitosis—all of these descendants necessarily living either in other groups or in descendant groups of approximately (with mixis) or precisely (without mixis—or mutation) the same composition. Second, they have also evolved to share equally in the reproductive opportunities afforded in their respective groups (this can be true even if there are rare circumstances when they do not share equally, as in meiotic drive or segregation distortion: for example, Dawkins 1989:235-36). As a consequence, once formed, the genome is (for the

most part in sexual organisms and entirely so in asexual ones) an indivisible unit with a unified function; the interests of all the genes in a genome are identical all the time (asexual organisms) or nearly all the time (sexual organisms). It is an aspect or outcome of this evolved identity that the organism has come to be an entity selected according to a singular principle: the maximizing of inclusive fitness via nepotism to either identical or only partially related descendants. Also as a consequence, the organism has acquired a lifetime: an ontogeny, a patterned reproductive period, and senescence. These are the kinds of "emergent" traits that group-selection may (at least in retrospect) be expected to yield. Some social anthropologists used to think that societies or cultures had such attributes, and were group-selected, but societies and cultures do not have ontogenies nor senescence, so to whatever extent they are group-selected, in the sense of their component individuals tending to have identical interests or having identical interests some of the time, the effect has to be extremely weak. Even during the momentary exception of reduction divisions and gamete formation in sexual (recombining) organisms, chromosomes and other subgroups of genes tend to remain intact, thus to be selected as units. Recombination is not differential reproduction; during oogenesis no chromosome or gene gains from ending up in a polar body, yet each allele has essentially an equal chance of appearing in the daughter cell destined to become an egg. A minuscule amount of differential reproduction occurs during this crucial less than one-millionth part of the life span, as a result of meiotic drive. To the extent that Mendelian expectations hold in inheritance, however, differential reproduction of alleles occurs because of the differential success of the groups (entire genomes) in which they find themselves, by virtue of differential success of the gametic and zygotic offspring of the individual organisms deriving from the zygotes containing those genomes (into which alternative genetic elements have roughly—or usually—an equal chance of getting). Despite implications to the contrary, (e.g., Rodseth 1990), inclusive fitness maximizing via nepotism is, therefore, not only a theory of the organism but *the* theory of the organism. It is not that the organism should be expected to have a general mechanism that instructs it to maximize inclusive fitness: the organism itself is that mechanism, and all of its component parts are continually modified and compromised in such fashions as to approximate that accomplishment. To understand fully any activity of groups or societies, we cannot avoid considering such basic facts about organisms as individuals.

The problem of ontogeny (development)—of the underlying mechanisms of inclusive-fitness-maximizing behavior—becomes the problem of precisely how the group-selected genes in the genome have evolved to cooperate in the production of the

organism (individual). The break in our understanding between the genes themselves and the organism and its functioning (hence, the interaction of genes and environment in yielding development and producing the organism), Waddington (1956) called "the great gap in biology." The implication is still largely correct (Alexander 1990). Understanding the general nature of evolution—long-term directional changes, and even the components of the evolutionary process and how they interact—may not be particularly difficult. Understanding the interaction of heredity and development well enough to assume an appropriate attitude with regard to the proximate background of behavior, however, is extremely difficult, and here the biological understanding of nonbiologists tends to break down. As a result, with dismayingly few exceptions, nonbiologists tend to become either naive enthusiasts who are wrong or naive skeptics who are wrong.

Having established that humans, as with other organisms, have indeed evolved to maximize inclusive fitness, biologists are now concentrating increasingly on the nature of underlying mechanisms of inclusive fitness maximizing behavior—hence, development, physiology, learning, the nature and consequences of evolved phenotypic plasticity, and particularly the mechanisms of kin recognition and nepotism (Hamilton 1964; Alexander 1977, 1979; Lumsden and Wilson 1981; Flinn and Alexander 1982; Holmes and Sherman 1983; Johnston 1982, 1985; Boyd and Richerson 1985; Tierney 1986; Fletcher and Michener 1987; West-Eberhard 1987; Symons 1987; Waldman et al. 1988; Cosmides and Tooby 1988; Tooby and Cosmides 1988; Barkow et al. 1991; Hepper, in press; for other references, see Alexander 1990, in press). The *existence* of such mechanisms can be established in any particular case merely through convincing evidence of evolved adaptive function. *Characterization* of the mechanism(s), on the other hand, requires knowledge of ontogenies, necessary and sufficient stimulus sequences, differential eases of learning, sensitive periods (Gould 1986; Hinde and Stevenson-Hinde 1973; Cosmides 1989), and ideally even geographic location and functional interdependence within the central nervous system and with respect to the minimal sensorimotor units (see Alexander 1969, for a detailed example from cricket song).

Efforts to focus attention on the proximate mechanisms of behavior in neural, developmental, and hereditary terms are by no means new. For example, it will be unfortunate if "Darwinian psychologists," who assert their mission to be the analysis of psychological mechanisms underlying evolved and adaptive behavior, repeat the futile arguments of the mid-century ethological "instinctivists." A half-century ago, evolved and adaptive behaviors were seen as plastic emanations from underlying innate or inherited instincts. The analysis of instincts was touted as the highest order of business for behaviorists. Controversies were prominent over how instincts might be inherited,

the degree to which their effects were genetically determined, how independent they were from one another, whether there was an underlying instinct for each behavior, how unlike learning instincts were, and what they really might be. These controversies culminated in an elaborate and specific definition of the hypothetical construct of "instinct" by a Cambridge Round Table Conference in 1949 (Thorpe 1950, 1951). This effort did not result in any "instinct" being identified or characterized in a way useful enough to last. Eventually the study of behavior and its underlying proximate causes seems to have continued much as it had before the special fuss about instincts.

In my opinion, most of the errors and oversimplifications in applying evolutionary theory, most of the arguments between biologists and nonbiologists on topics such as sociality and morality, and most of the seeming disagreements among investigators over mechanisms result from simplistic views of development, hence, of underlying mechanisms of behavior. A crucial problem in human self-understanding is incomplete knowledge of the nature of motivation, both in general and by individual actors concerning their own lives. It is not satisfactory merely to apply terms indicating ignorance of development, such as "innate." Nor is it helpful to invoke plasticity vaguely, referring to a behavior simply as "learned" and pretending that the problem has been solved. It is not sufficient to say that any behavior is "genetic," or to assume that individual genes determine any behavior whatsoever. Stages and events in the developing organism are inevitably epigenetic—not only influenced by the genome as a whole but controlled by feedback from the developing phenotype as a whole:

> . . . the idea of genes containing 'blueprints' for behaviors is probably false. I will emphasize . . . the concept of nervous system development as an epigenetic phenomenon in which, once set into motion by the entire genome, one event simply leads to another. A functional nervous system depends on the astonishing ability of developing neurons to select among thousands of candidate targets precisely the correct cell with which to establish a connection. But this precision is best viewed as a developmental consequence of the two cells' being in a particular place at a particular time rather than of genes' specifying the formation of neural circuitry . . . specific genes do not control specific events in normal nervous system development. There are not one-to-one correspondences between genes and neural structures, but rather many-to-many relationships, with each developmental stage guided by epigenetic information . . . (Tierney 1986:341-42)

Suppose that there is a gamut from phenotypic attributes that represent many-to-many genetic and ontogenetic relationships to another extreme in which the genome (organism, phenotype) "allows" control of a difference between phenotypes or a change

in a phenotype by a single gene's effect such that it appears that "a gene" controls "a" behavioral variation. Nevertheless, any such gene's effect, and its activity or non-activity, is under the control of the organism as a whole, therefore epigenetic, and results from the relationship between the internal and external environment and the genome as a whole, rather than being in any (alternative) sense genetic or particulate. Eventually we must understand how the flexibility (through responses to environmental variations) represented by the overall phenotype itself, and particular kinds of flexibility such as learning, have originated, and how they have been formed, adjusted, biased, and elaborated by natural selection (Johnston 1982, 1985; Tierney 1986; West-Eberhard 1987; Alexander 1979, 1990). We must unravel the ways in which the unitary phenomenon called the organism, with its cooperative collective of genes evolved toward the singular evolutionary function of inclusive fitness maximizing, has been created from what Dobzhansky (1961) called "the particulateness of heredity." We must understand the unity of development in environments rendered modular by the independence and multiplicity of the different "Hostile Forces of Nature" (Darwin 1859). We must develop concepts of different kinds of epigenetic preprogramming to replace notions of inherited, genetic, genetically determined, "innate," or "instinctive" behavior (or templates or substrates). The last two terms are simply vague and useless, but they lead us to the others, which are wrong, and promote erroneous thinking about the interactions of heredity and development and the nature of phenotypic plasticity. The genes may cooperate to produce a singular organism, but natural selection impinges on the organism as a result of reproductive failures and mortality from many different sources: climate, weather, food shortages, parasites, diseases, predators, and mate shortages (all such problems, of course, mediated by conspecific competitors). As a consequence partly of the multiplicity of selective forces, the organism evolves traits (and underlying mechanisms of behavior). But the traits of an individual organism, however separate they may seem, are not independent of one another, as is well understood by any biologist who has tried to interrupt the "stream" of activities (or the continuity of morphological or physiological features) of an organism to itemize its traits in an ethogram or to maximize their numbers to, say, construct the most likely phylogenetic tree. No trait of an organism is maximized in its own particular function because all traits are part of a compromise in which the singular function of inclusive fitness maximizing remains as the perpetual combined effect of natural selection on the organism. Evolutionary compromises within the evolving organism as a result of conflicts among the "idealizing" of different functions are parliaments not so much in the sense of conflicting interests as in the sense of coordinations of extremely complex

507

programs of effort (and possibly of differences in information among agreeing parties—or parts). If this unity is disrupted by any gene behaving other than as group-selected in its genome, the consequence will nearly always be disaster; the exceptions are alleles that show meiotic drive (e.g., reduce their chances of going into a polar body) *without* showing accompanying deleterious effects on the organism's phenotype.

In what sense is plasticity—for example, learning—primitive or derived? The phenotype as such—the original and general manifestation of flexibility—is ancient, so that any of its current features in any organism are likely to represent successive waves of derivation. It does not follow that plasticity *per se* is invariably derived from non-plasticity, or greater plasticity from lesser plasticity. Learning is not always derived from nonlearning, or greater flexibility in learning from less (Johnston 1985; Tierney 1986). For example, plasticity or flexibility *per se* can be seen as either adaptive or as "noise." Consider a communicative signal. At first it may be variable, with the variability resulting from the primitiveness of the signal, perhaps best termed imprecision, and having little reproductive significance because all variants of the signal still work. If, however, some parts of the variability come to cause confusion with, say, signals by other species new to the region, then variability (plasticity) is likely to be reduced accordingly. Different parts of the range of variability in a signal may begin to have different levels of significance in different life circumstances for the species (say, more intense signals are more useful for pair formation because they attract mates from a greater distance, and less intense signals for courtship because they lessen interference by attracted competitors). Now the variability may be compartmentalized, producing "nodes" of functionality within the original range of variation, with the eventual effect of reducing observable "plasticity." One might expect that, for any particular trait, "waves" of different kinds and degrees of plasticity would occur during evolution, with perhaps repeated tendencies to modularize after "rushes" of plasticity or variability had occurred—for example, when a new habitat is entered or a new function appears. Thus, in the earliest stages of song evolution in crickets, a reasonably wide spectrum of frequencies was surely produced. While such frequency spectra might appropriately be described as physically "complex," functionally this complexity would be an aspect of rudimentariness and primitive variability rather than evolved precision. The ability of most modern crickets to produce a narrow range of frequencies (essentially a sine wave) has to be regarded as derived or a specialization. Similarly, the precise rates and rhythms within the songs of modern crickets are surely not primitive but derived from less precise, more variable primitive ancestral efforts. When wings (secondarily used to produce song in crickets) evolved in any organism, they may have passed through

early stages in which they were modified as gliding organs (probably after serving some entirely different function at an earlier stage). When actual propelling features appeared, a new wave of selective change would take place in a slightly different direction, probably leading to a new "rush" of variability. As different flying techniques or advantages appeared, further changes would occur in different directions, and in succession, within any genetic line. In modern forms wings have evolved several times and are used for not only flying and signal production (e.g., crickets, katydids, butterflies, birds), but have evolved many other functions such as fighting (e.g., ducks, geese), food and mating lures (katydids, crickets), and grasping and holding devices during mating (scorpion flies). Each time a new effect with reproductive significance appears, selection causes changes in one or more different directions, sometimes seeming to alter the wings in a general overall fashion, sometimes seeming to modularize their structure and function. In many cases plasticity will seem to be reduced following a stage in which multiple uses occurred rudimentarily because of what would be perceived as a phenotypic variability or plasticity.

It seems reasonable to hypothesize that the kinds of behavioral flexibility we call learning have evolved in fashions similar to those of other traits; there is no reason to restrict ourselves to hidden features of the human mind or CNS when trying to understand such general features of the organism as the underlying physiological and psychological mechanisms of behavior. In the above example, wings would be parallel to learning ability, the changes in wings as a result of selection favoring different functions paralleling modularization or multiple biasing of learning to favor its uses in different life circumstances. Despite any "domain-specific" aspect in which wings (or learning) evolved to serve some particular function, the "domain-general" phenomenon (of wings or learning) would remain. "Domain-specificity" is necessarily a relative term, so that to use evidence of it to deny "domain-generality" seems futile. We ought not to let the extreme "blank slate" arguments about learning (Skinner 1965) distract us from this realization. It can also be seen from this example that the question whether plasticity or canalization (learning or some alternative to it) is genetically or developmentally more expensive is not easily settled (Johnston 1985; Tierney 1986).

Some traits of organisms are identifiable, or give the appearance of being partially independent, because of the modularity of the selective environment. All such traits, however, cannot be explained in this way. Consider a gasoline engine. It is designed for a singular function, to create maximum torque at a certain speed on a particular shaft. Nevertheless, it will have different "traits" just as does the organism: cylinders, pistons, valves, spark producers, distributors, timers, cooling devices, intake

and exhaust systems, crankshaft, fuel tank, etc. Some of this modularity occurs because of previously perceived efficiency by the designers. Worn-out parts are more easily replaced if they are made separately and fastened together. Some components of an engine are more easily constructed by creating separate parts and fastening them together. Even excepting these examples, "traits" will still exist, just as organisms have separate circulatory, filtering, locomotory, and alimentary devices. Part of the evident particulateness of the organism as well as the engine therefore is due to efficiency of specialization of parts. But a significant part of organismic phenotypic modularity is surely caused by the particulateness of the external environment.

What are the reasons for heredity remaining particulate? Because of the tendency of "genes" to live in variously tight large and small groups (supergenes, chromosomes, and other kinds of "linkages"), yet to remain incompletely "congealed" even in secondarily apomictic (non-recombining) forms, we might speculate that the manner by which the genetic materials are changed is partly responsible. That is, perhaps heredity remains particulate at least partly because radiation and other mutagens tend to influence small regions in the genetic material, resulting in what are called "point" mutations. Collections of such "point" mutations are generally responsible for individual organisms differing genetically in particular amounts and ways, and therefore lead to the differential reproduction that creates the patterns of life. Nevertheless, when "point" mutations occur on chromosomes, except for rare breaks in the chromosome (and the resulting inversions and crossovers) the chromosome as a whole is altered, and gene changes are either saved or lost primarily because whole chromosomes are saved or lost. To this extent the genome has in fact congealed and heredity is not so particulate. Explaining the particulateness of heredity is different from accounting for the retention of sexuality; recombination is facilitated by the particulateness of heredity, but its advantages may not be responsible for the retention of particulateness.

To go beyond avoiding the oversimplifications and mistakes we already recognize, in discussing underlying mechanisms of behavior (thinking of learning as simply a "blank slate" or unbiased, undirected plasticity; referring to behaviors with cryptic ontogenies as "inherited"; conjuring up misleadingly precise definitions of hypothetical underlying mechanisms as "instincts" or "psychological mechanisms," visualizing the central nervous system as a single-layered conglomerate of innate, specialized, separate, and independent mechanisms for particular behaviors), is a monumental task in which behaviorists will eventually have to depend heavily on developmental neurophysiologists. In turn, developmental neurophysiologists will have to depend on evolutionary biologists to understand how the various phenomena they analyze can be under-

stood as "mechanisms." Part of the analysis cannot be completed without knowing what each mechanism is evolved to accomplish—therefore, knowing the evolved life functions or adaptations of the organism as a whole. Social scientists and philosophers will have to keep up with at least the bare bones of the arguments.

Extrapolating from the best-studied phenotypic attributes in the most familiar organisms, mechanisms evolve which tend to yield particular behaviors in particular environments. To understand them behavior must be studied first and the mechanisms only later. Nothing requires (or allows!) that underlying mechanisms of behavior be identified prior to understanding what they have evolved to do, and nothing requires that any particular mechanisms—"epigenetic rules"—be unchangeable.

As for virtually all aspects of all phenotypes in all organisms, nothing about the human phenotype involving morality either has been identified or need be identified as innate, instinctive, inherited, or unchangeable. It is a fundamental misunderstanding of inheritance and development to believe that such a requirement is necessary for a biological discussion, or that it is assumed in every discussion by a biologist about the evolution of behavioral tendencies and abilities such as those involved in morality. Both genes and environment contribute to every aspect of the phenotype, and the evolutionary significance of the concept of phenotype in all its ontogenetic or life stages is plasticity, in the interests of maximizing inclusive fitness in the environments of history. Genomes, which are by definition nonplastic and unchangeable (except by mutation and recombination), do not occur (any longer?) naked in the environment (that is, without phenotypes). Presumably naked genes have been universally less capable than organisms of adjusting appropriately to environmental contingencies, and on this account less persistent. Ontogenies are increasingly complex systems of feedbacks involving the the entire genome and the internal and external environment of the organism, turning genes on and off and modifying their actual effects. It may be possible to conceive of ways that certain phenotypic effects do not represent plasticity, but in general phenotypes evolve as cushions of plasticity interposed by selection between the genome and the environment. Even when a phenotype consists of what seems to the observer an entirely rigid structure or unalterable activity, the evolutionary (and ontogenetic) production of that expression represented a change in the previous "organism," thus flexibility in that sense, and as well in the sense of susceptibility to further change when selective forces change again. Once phenotypes evolve, gene frequencies change because of the particular structures genes produce in particular environments, and the significance of the structures (phenotypes) is that they alter the way selection acts on gene differences. The different kinds of plasticity that represent the phenotype, includ-

ing learning, are ways of providing reproductively appropriate responses to varying forces of natural selection but not ways of escaping them entirely.

It seems fallacious to argue that we are evolved beings whose every attribute has been produced by an evolution guided primarily by differential reproduction, and then to follow that argument with the assertion that nevertheless we have somehow magically and more or less suddenly escaped the underlying mechanisms that have resulted from this continual, inexorable process of natural selection that is our heritage (regardless whether or not the evolved mechanisms happen to be maximizing reproduction in current environments). This is not the way, for example, that long-term effects of culture would be expected to have changed us. If the underlying mechanisms of our behavior do not persist unchanged, they surely have been changing so as to render our behavior in the cultural situations of recent history more reproductive, not less. The challenge is to describe how the underlying mechanisms of reproductive maximization are working in modern society, not to interpret all of modern human behavior as if the mechanisms had somehow disappeared. Culture is ancient, and the learning capacities on which it is based are even more ancient. As culture developed through the ages, human learning either served individual reproduction or it did not; we can be sure the parts that did not have been damped by natural selection. Regardless of the speed of cultural change, there has been ample time for much alteration of learning tendencies and capacities since culture began. I am not aware of evidence that learning capacities have been diminished, in any general sense or measurable degree, since the origins of human culture. They may, however, have been altered dramatically in the service of individual reproduction. In this there is no suggestion of escape from the evolved mechanisms of our behavior. Any semblance of "escape" is most likely as a result of the particular kind of dramatic, rapid, and potentially progressive (sequential) learning represented by new understanding of the evolutionary, reproductive backgrounds of our behavior.

Particular examples of misunderstanding of the relationship between evolved ontogenies and the nature of human motivation, decision-making, and choice may be instructive. Thus, some authors (e.g., Slobodkin, this volume) suggest that if organisms show nepotism to relatives solely as a result of having been reared with them (rather than, say, through having evolved mechanisms for identifying strangers who are kin), this means that cumulative genetic change through natural selection is not responsible for the effect. Such a view implies that social learning is not a consequence of evolution and is not designed to cause reproductive behavior in organisms living in the environments of history; that learning particular things from association at particular

times and in particular circumstances cannot comprise a mechanism of inclusive fitness maximizing. We know this view is false, since learning about helping relatives and . avoiding inbreeding is patterned in the ways denied by it, and they are patterned differently in group living and solitary forms precisely in ways predicted by evolutionary considerations (Alexander 1979, 1990, in press). Moreover, the inadequacy of such views is not merely academic: knowledge of sensitive periods and learning biases have implications of great significance in resolving human suffering resulting, for example, from homicide, incest, and child abuse (Daly and Wilson 1981, 1988; see also, Alexander 1987, 1989). Knowledge of learning experiences that are not conscious, hence not remembered, may have remained trivial in terms of conscious reflection (that is, failed during evolution to become conscious, and as a result failed to *seem* important) because in the environments of history they did not typically go awry without the intervention of consciousness. Current social environments, however, may cause such experiences to go offtrack considerably more frequently than is necessary or desirable, making their conscious understanding exceedingly valuable. It is not trivial that associative learning, as a mechanism of kin recognition, serves the interests of all the genes in the genome equally, therefore accords with the earlier description of the organism as evolved by group selection of genes; some other proposed (but not demonstrated) evolved mechanisms of kin recognition do not (Alexander 1990, in press).

Patricia Williams (this volume) argues that theories of prescriptive evolved ethics are internally contradictory because, to be ethical, organisms require freedom from external and internal coercion. She seems to mean that an organism evolved to maximize its inclusive fitness is coerced by this aspect of its makeup, so that either ethical behavior is independent of evolution or evolved beings cannot be ethical—as she puts it they cannot legitimately be blamed, praised, and held responsible for their actions. For this argument, however, she must also assume that the existence of rules, . and of threats about the consequences of breaking them, is independent of the existence of moral and ethical behavior. It seems to me that her view also requires that many decisions of humans (all of those in the realm of ethics and morality) are made in the absence of potential costs or independent of potential costs that could, in a broad view of the concept, be viewed as "coercive"—or at least in the absence of any kind of knowledge (conscious or nonconscious) of such costs. I would argue, to the contrary, that every act of every organism involves potential costs and potential benefits, and some kind of cost-benefit analysis, so that in this sense every human decision is "coerced"—including the formation of conscience itself. To take this position one need not argue (and I never have argued) that any particular ethical or moral rules will ever

be discovered to follow naturally or inevitably or at all from evolutionary facts. Rather, I would argue that acts *considered* moral or ethical are typically acts that result in a perceived temporary net cost, or risk, to the actor, in the process of giving benefits to others, particularly when there is evidence that the actor did not consciously calculate the cost or risk and determine that the likelihood of eventual overcompensating return actually made the investment worthwhile (Alexander 1987). To me, the interesting or difficult question is *why* humans tend to regard only acts of this nature moral or ethical, and what are the consequences of this tendency not only for the active investigation of motivation in social behavior but for the prospects of eventual widespread acceptance and understanding of organic evolution as the causal sequence giving rise to humanity and its traits and tendencies? Returning to Patricia Williams's comments about praise and blame, I think we praise people for following the rules of society even if they do it as a result of recognizing that it will be even more expensive to break them, and that we tend to blame them for failing to follow the rules in nearly all cases except when they are incapable of comprehending them. In any case, the existence of rules, and of admonitions concerning them, alters people's behavior; this must be a consequence of evolution and certainly is in no way contrary to it.

Even if I have not erred, and Patricia Williams really is using the term "coercion" to mean essentially any potential costs that might modify actions, I nevertheless would not wish to imply that in the everyday sense of the word any and all costs represent coercion. It seems that, as with any rule-based system, to give the concept of coercion a useful meaning one has to be willing to specify degrees of restraint, hindrance, or compulsion that are appropriately termed "force" and others that are not, and more specifically kinds or degrees of force that are considered legitimate and illegitimate. Thus, morality is not usually seen as requiring that everyone entirely cease serving his or her own interests and serve only those of others. Rather, what is required is that the effort to serve one's own interests not exceed certain limits in directions that interfere with others serving their own interests. The appropriate limits, I think, tend to be set according to some kind of opinion—of a majority or a power structure—and such opinion tends to become highly dependent on precedents. What constitutes unacceptable coercion (or imposed costs) of particular acts also tends to depend upon precedents. Precedents lead to expectations, and the parading of precedents reinforces this tendency. We often speak of interfering with legitimate expectations, and I would suppose that such considerations often enter into people's notions of moral and ethical behavior. Some degrees and kinds of "coercion," in the form of restraints or hindrances or costs, thus seem to exist for essentially all courses of action.

George C. Williams (1989, this volume), author of the bleak statement quoted at the outset, attempts to solve the problem of evolution and morality by following Thomas Huxley (1896, reprinted in 1909) in declaring that natural selection is evil and morality combats it. According to this view, actually resembling that of Patricia Williams, to serve the interests of morality we presumably act intentionally counter to our evolutionary interests. Unfortunately, this view does not explain what possible impetus we might have to pursue such a course—or to be moral. It provides no explanation for the striking human tendency to generate moral systems, discuss morality endlessly, and adhere to moral rules in particular circumstances. It does not explain why humans tend widely to identify as immoral actions within their groups that they nevertheless regard as highly moral between groups. Ultimately, George Williams's view would have humans no less amoral than any other form of life. He seems to refer only to the fact that evolution tends to serve individual interests and to pass over the question of how individual interests may be served by serving the interests of the individual's group, however composed for particular and different questions pertinent to that proposition. His analysis ignores that cooperation, which to many seems antithetical to the service of individual interests, is actually consistent with natural selection, and it seems to deny that cooperative behavior which facilitates reproduction can ever be moral. In this ignoring and denying, it also thwarts any likelihood of understanding how tendencies to cooperate, and indeed to generate and espouse the kind of view Huxley and Williams put forth, could have come about.

The Evolutionary View of Morality as Contractarian

Because selection is primarily effective at and below the individual level, it is reasonable to expect concepts and practices pertaining to morality—as with all other aspects of the phenotypes of living forms—to be designed so as to yield reproductive (genetic) gains to the individuals exhibiting them, at least in the environments of history. To put such an approach into practice it is necessary to understand the life interests—therefore the life patterns—of humans as outcomes of an evolution guided principally by natural selection. This also means understanding theoretical arguments from the science of biology about such things as somatic effort (and ontogenies), reproductive effort, and senescence (Williams 1957, 1966; Hamilton 1966; Alexander 1987).

As with most organisms, humans serve their interests as individuals by interacting with others of their species, even if they interact in some ways unique to the

human species. Their interactions are not merely competitive but also cooperative. They give and receive benefits from one another. Sometimes they compete by cooperating. Both their cooperation and their competition can be extremely complex. We all know these things, although it often seems that there is an undue emphasis on cooperation in a way that refuses to acknowledge its competitive aspects, speaking of cooperation as a replacement for competition and talking only of the "brighter side" of human nature. If humans had evolved to cooperate only to avoid nonhuman hostile forces, there might be no reasonable argument that self-deception is involved in efforts . to argue for cooperation *as an alternative to* competition (rather than as a form of it). Since Darwin, however, we have known that this is not likely to be the case; current arguments that the human intellect evolved as a social tool, moreover, seem to lend potent fuel to the opposite argument, that human social cooperation may literally have evolved as a method of intraspecific competition (Humphrey 1976; Alexander 1989, 1991).

In humans, uniquely, intraspecific competition has become dominated by intergroup competition facilitated by cooperation in large numbers of coalitions that overlap in indefinitely large and complex patterns, both vertically (between generations) and horizontally (within generations). Nepotism to both descendant and collateral relatives, and reciprocity involving both relatives and nonrelatives are the means of dispensing and withholding benefits in patterns appropriate to the serving of evolved interests. Such beneficence in the contexts of evolved nepotism and reciprocity—whether in evolutionary terms appropriately or inappropriately directed today—is actually what has been termed the altruism of morality; selfishness, in the sense of egoistic behavior, is expected to be designed in the interests of creating a phenotype maximally capable of the beneficence of nepotism. In the service of inclusive fitness maximizing, the organism may thus be described as evolved to be the most effective possible nepotist (Alexander 1979).

Through indirect reciprocity, even beneficence to nonrelatives unable or unlikely to reciprocate may be favored because it can lead to the establishment of reputation that benefits the beneficent individual with respect to later choices of partners in reciprocity (Trivers 1971; Alexander 1979, 1987). Centrality of indirect reciprocity in modern human social systems creates situations in which reputation depends significantly on (1) evidence of general acceptance of, and adherence to, sets of rules and (2) the presence of conscience as a vehicle to such behavior, as opposed to simpler forms of indirect reciprocity depending on beneficent individuals simply repeating the same kinds of altruism in the same kinds of situations in which they have previously been observed.

Under this model reputation may become so important that for some people in some circumstances indiscriminately dispensed beneficence can be favored (Alexander 1987).

Thus, along with others, I have argued for a contractarian view of morality, thoroughly understandable only through knowing, at the outset, the life goals or interests likely to be produced by organic evolution, but relative in the sense of being modifiable to fit particular social situations and including enormously complex patterns of cooperative and competitive interactions. I argued that intragroup cooperativeness is designed evolutionarily to serve intergroup competitiveness. The universality of intergroup competitiveness and the seeming altruism of within-group morality, I maintained, are together responsible for morality seeming erroneously to be a group phenomenon that serves the interests of the group contrary to the interests of the individual; they represent the vehicle to understanding (1) the correlation between the existence and the intensity of within-group amity and the existence and intensity of between-group enmity and (2) the failure of humanity to accomplish the cultural derivation of a single cooperative group encompassing all of humanity with a common set of moral rules—a goal otherwise seeming to many philosophers and moralists as the logical but perplexingly unachieved extension of within-group amity. I have also argued that the significance of indirect reciprocity has to do not only with rules but with intent, that systems of indirect reciprocity lead to avoidance of selfishness as well as positive acts of (at least temporary) altruism, that we are evolved to appear to be honest and altruistic, and that "We use motivation and honesty in one circumstance to predict actions in others." (Alexander 1987:96, and Table 2.5). I have argued consistently that, even though evolution and natural selection directly or indirectly underlie all our actions and motivations, conscious understanding of our evolutionary background does not provide easy ways either to interpret or to justify opinions about right and wrong in particular situations. One reason is that right and wrong are group decisions, sometimes by majority opinion, sometimes by a power structure that may represent a minority, and sometimes only by unanimity. With regard to human behavior selection has not worked at such levels; as a result individual opinions, however learned, are unlikely to represent group compromises, and, when they reflect conflicts of interest, may even conflict ludicrously with them.

Among the many publications about morality in the last five years or so, a contractarian view seems to have gained ground. Moral behavior is evidently a result of some kind(s) of contractual agreements among participants in sociality. Moral contracts are somehow less formal than the more familiar legal contracts—maybe even less explicit—but they are there, as contracts, all the same. Rawls (1971), Gautier

(1990), and Gibbard (1990) are examples of moral philosophers who view moral behavior as a result of contractual behavior (see also Axelrod 1984, 1986; Frank 1988). Legal contracts might have arisen out of the history of moral, and perhaps other kinds of verbal or somehow understood and accepted, contracts by which society operated, and that written language was one reason legal contracts in their present forms became possible. Written language necessarily enhanced the possibility of unequivocal precedence, and the concept of precedence enhances the basis for rules. Unlike moral contracts, however, legal contracts are always *consciously* entered into and developed, and they are not always regarded as moral.

Contracts are made between people who have real or potential conflicts of interest. There is no other reason for making them. There is no reason for a contract between people whose interests are identical and will remain so. Among sexually reproducing organisms, in which every generational link halves the likelihood of any particular gene being present in any two related individuals, the (evolutionary) interests of individuals will rarely, and then only temporarily, overlap completely (Hamilton 1964; Alexander 1979, 1987). The implication is that through understanding the evolutionary significance of the human organism, and the nature and individuality of its evolved interests, we may derive useful insights into human concerns about morality.

Moral behavior refers explicitly to the *rights* of others. It represents how far anyone can go without being judged unfair, unjust, wrong—without being judged immoral. How far anyone can go means, presumably, how far one can go in serving his own interests. In other words, the rules of society—whatever they may be or whatever kind they may be—are designed to control the tendencies of individuals to serve their own interests in various ways, presumably in ways that have minimally deleterious effects on those who make and maintain the rules (sometimes virtually all of society), and sometimes have downright positive effects on those other than the individuals being inhibited.

In my experience people do not regard themselves as serving only their own interests—not even as serving only the interests of themselves and their own family. How people think about motivation, and that how they think about how motivation is likely to be altered by statements about motivation, represents a principal difficulty in securing widespread acceptance of evolution and natural selection, and an evolutionary approach to human behavior. That sophisticated writers such as Frank (1988) and Gibbard (1990), who attempt evolutionary analyses of aspects of human behavior, can nevertheless (in my view) underplay human activities and tendencies prominent in the daily news (that is, selfish and exploitative behaviors: see below) illustrates the potency

of this difficulty. I believe that it also accounts in part for the curious tendencies of authors such as Hayles (1990) to attempt to relate science and human nature by skipping from the physical sciences to moral philosophy or literary criticism without significant mention of the science of biology or the facts of evolution.

The Problem of Intent

Earlier I suggested that the most paradoxical problem in applying evolution to human behavior—of analyzing the evolution of concepts and activities related to morality—is that of understanding motivation. People do not see themselves as designed to maximize inclusive fitness. They do not think of their activities as serving only reproduction. They tend to be hostile to any concept or discipline that seems to rely upon this kind of reduction. This intuitive rejection, moreover, is only part of the paradox. Consider expression of the emotions, much of which is tied to intent and the communication or expression of intent to others. From an evolutionist's viewpoint it is reasonable to hypothesize that initially expressions of physiological changes that incidentally but accurately signaled intent—or imminent action—were used by other individuals to their own benefit. To the extent that this occurred, organisms surely began to evolve so as to alter externally perceptible evidences of intent to serve their own interests rather than those of observers, at least when the two sets of interests conflicted. Assuming that this is a reasonable view of the evolutionary background of current expressions of intent through the emotions (Alexander 1989), it is extremely perplexing that humans are to a large extent imperfectly or incompletely in control of expression of the emotions. Everyone blushes sometimes when not wanting to. Everyone has lost his or her temper and regretted it. Everyone has suffered embarrassment because some indication of sexual excitement was not concealed. Probably no one regards him or her self as always in complete control of the emotions (that is, in complete *conscious* control).

I doubt that many evolutionarily oriented behaviorists would accept that lapses of the sorts I have just described are simply evidence of failures of natural selection. Why should such things happen if expressions of the emotions—as indicators of desires or intent—should exist to serve the interests of the individuals expressing their emotions? Why should humans have evolved so as not to understand their own motivations, and not to be able to control even the evidence of the motivations that they believe they do understand? It is difficult to imagine that we are here dealing merely

with inevitable concomitant costs of other tendencies.

Biologists typically do not ask about the intent of nonhumans, the most obvious exceptions being our closest relatives such as chimpanzees, bonobos, orangutans, and gorillas. Nonbiologists concerned with establishing the connections between natural selection and human behavior not only lack the biologist's background in examining endless evolved features without being concerned with intent, but rather are more likely to be preoccupied with intent from the beginning. This preoccupation does not derive solely from the fact that anyone who studies or thinks about humans must constantly consider intent. It also arises out of the difference between scientific investigations and the pursuits of the humanities. Science is a process of accumulating knowledge, with the aim of approaching irrefutability or undeniability—of discovering factual information of general and uniform applicability. Work in the humanities—with some of the most complex and difficult-to-understand human activities—seems, in contrast, primarily a search for meaning (Alexander 1989, 1991). The inevitable and perpetual individuality of interests among humans (itself a product of evolution through sexual recombination, leading to genetic individuality—see Alexander 1987, and Williams's quote above) indicates that meanings will tend to be different for different individuals. That generalities of meaning also exist does not alter the fact that the emphasis of the humanities is on human endeavors in which meanings will always tend to be diverse, and often individual. Thus, when I state my personal intent in a given matter, to the extent that I am being honest I am describing some aspect of the individuality of meaning in that matter. This distinction is surely responsible for some controversies currently prominent about the nature of reality and the importance of individual viewpoints, extending even into the difficulties of discussing cultural and other human variations and judging the outcomes of social interactions, the meaning of which is seen differently by different participants.

Expressions of intent, and other uses of intent, as with other traits, have presumably been shaped by natural selection. As already suggested, many aspects of communication, such as expression of the emotions, probably evolved originally not because they transmitted any honest information at all but rather because they were useful coverups for honest information originally produced only incidentally and as such detrimental to the interests of those producing it because of its use by others (Alexander 1989). The notion of communication as simply a system of transmitting honest information to social associates has long been dead (Otte 1974; Lloyd 1977; Dawkins and Krebs 1978; Alexander 1987). Whether or not it includes honest information, communication is also likely to involve deception and misinformation. It will have to be

analyzed in terms of the degrees and kinds of overlaps of interests among communicators, and such analysis will depend upon understanding histories of genetic relatedness and the means available to humans of recognizing and responding to patterns of common and differing reproductive interests.

Frank (1988), an economist, dealt extensively with the problem of intent in *Passions within Reason*. His initial thesis seemed to be that a self-interest model of human behavior will not account for morality—for some easily observed instances of beneficence or altruism such as the leaving of tips at restaurants visited but once, following one's conscience, or returning found money. His arguments about costs and benefits depend on whether or not there is any possibility of being found out (that is, chances of imprisonment or effects on reputation), and although he declares that detection is impossible in some of his examples, it is doubtful that one can ever be sure of this. I would guess that the prisons are filled with individuals who at one point or another were certain they would not be caught.

Frank attempts to show that many acts of beneficence are indeed unmitigated altruism—net-cost acts for the actor. Then he argues that the reason such acts are undertaken is that the actor gains sufficient benefit from having previously committed himself to such acts that, should the occasion for them arise, he will actually to reap a net social benefit overall. This benefit comes both from the responses of social interactants to the evidence of the existence of the commitment, and to the readiness of the actor to act in ways seemingly contrary to his own interests without having to ponder the question (Alexander 1987). In other words, just as it pays not to dwell consciously on the rhythm of the heartbeat or how fast to breathe, it sometimes pays to conduct social acts without prior conscious calculation, even if as a result of foregoing conscious reflection occasional mistakes occur.

Only at the very end of his book does Frank acknowledge what the reader has long before begun to suspect: that, after all, his arguments do indeed fit a self-interest model—just not a *conscious* pursuit of self-interest. *Commitment* actually becomes his term for nonconscious pursuit of self-interest, which pays both because of the effect of commitment on other social interactants and because most of the time commitment leads to net-benefit acts. Frank believes he has explained why *some* acts caused by commitment may be a net cost to the actor. When what he had originally described as altruistic acts are evolutionary accidents, they are not explainable directly from natural selection. Such acts are often explainable, however, in the same way that, say, senescence is explainable from natural selection—indirectly—in the case of senescence as inevitable concomitants (or pleiotropic effects) of otherwise beneficial genes (Williams

1957). It is appropriate and useful to point out the inevitable concomitant costs of evolutionary designs, and that there are always such costs; it is not the case, however, that biologists have been operating as if such costs did not exist, or that they deny a self-interest model of human behavior.

In the end Frank's analysis, although extremely enlightening from its ease of reading and wealth of examples, seems not to differ substantively from others previously generated around the concept of indirect reciprocity (Trivers 1971; Alexander 1987). Self-interest as inclusive fitness maximizing remains the central evolved goal of the human organism, and the concept of commitment simply represents a particular kind of contract with society, or some part of it. The difference between Frank's self-interest and commitment models is not whose interests are being favored but whether or not costs and benefits are being calculated consciously.

Similarly, Simon (1990) argues that there is much net-cost altruism among humans and that it is induced by social learning that has a net benefit but entails costs because it sometimes leads to mistakes. He points out that it is often to the advantage of others (or to "society") to induce such mistakes, which he calls a "tax" imposed by society "on the gross benefits gained by individuals" from what he calls "docility," which translates as "the human tendency to learn from others." (p. 1665; also, Alexander 1987). He also notes that what he calls "docile" acceptance of others as authorities for phenomena of great significance in our lives takes place because in complex human society it is often prohibitively expensive to seek the answers for ourselves. Costs resulting from the wrong things being learned have long been understood as an inevitable concomitant of learning. That there are social benefits from being committed to social beneficence, that even indiscriminate altruism can be beneficial to the actor, and that people exert effort to get others to behave in ways beneficial to others and deleterious to themselves has also been argued extensively (e.g., Alexander 1987, especially pp. 96-126). Simon, as with Frank, seems primarily to have rediscovered indirect reciprocity and perhaps not to be aware that all acts of all organisms involve costs.

"Docility" and "commitment" are thus aspects of indirect reciprocity, with the costs to foregoing immediate benefits for long-term benefits on average over-compensated by those long-term benefits. In other words, both Frank (1988) and Simon (1990) are suggesting that you cannot accept learning and following the rules, and develop your tendencies to their fullest in regard to returns from this procedure, without incurring costs from occasional mistakes of beneficence that would not otherwise have been necessary in the particular circumstances in which they occurred.

Neither Frank nor Simon provides an answer to the next question: Why are we not in complete control of the use of the expression of our emotions to manipulate others? Earlier, though, Frank says, "A blush may reveal a lie and cause great embarrassment at the moment, but in circumstances that require trust, there can be great advantage in being known as a blusher."

Blushing is not restricted to situations involving lies. Sometimes it occurs when we simply are thinking thoughts we do not wish revealed. Are we evolved to give the impression (sometimes true, sometimes not) that in certain situations we cannot control our emotions therefore cannot accurately convey our intent or regulate it? Or else that we have incomplete intent (are undecided) but do not wish to convey (reveal) that fact, though we are indeed revealing it (against our wishes) and presumably are thereby benefiting (at least in the long run)? Are we designed to convey incompleteness or imperfection of intent (or motivation)?

I have argued (Alexander 1989) that we may have evolved incomplete control of the emotions to dispel the notion that we are in complete control of our emotions in particular situations. This hypothesis implies that incomplete control of the emotions— or unwanted expression of them such as blushing, uncontrollable anger, or uncontrollable sexual excitement—occurs in interactions with others of great importance to us, especially when we are establishing, cementing, or furthering the relationship, and much less or not at all in casual or brief interactions. Without such lapses we appear as "cold fish" who have our emotions entirely under control, or as uncommitting individuals who are not undergoing the changes and indecisions involved in establishing an expensive social investment, therefore that the current involvement is not particularly important to us. The costs of occasional mistakes from not employing our conscious abilities in all social encounters may sometimes be high, but the cost of being discovered to be insincere is usually enormous.

The vast majority of expensive human social interactions across history have probably been nepotistic, and involved extensive commitment or investment to either relatives or spouses. From this it could also be argued that some social interactions used extensively by Frank, such as tipping in a strange restaurant—in other words actions involving the question of beneficence toward strangers that will not again be encountered—are to a large extent novel or actually apply to (or are incompletely associated with) intergroup interactions, in which (sometimes) short-changing and other versions of what Sahlins (1965) called "negative reciprocity" are not only expected but admired.

Incompleteness of control of expression of the emotions is not part of the

general public's problem with evolution: only the serious student of human behavior in relation to evolution is likely to expose it and worry about it. If critics of evolutionary approaches to human evolution had been aware of it, they surely would have exploited it as a major part of the reason that an evolutionary view of human behavior does not seem to "make sense" and has not spread to include a much larger portion of the world population of educated humans.

Consider this set of arguments:

Human society cannot operate without rules because too many people—acting as both individuals and groups—will go "too far" in serving their own interests; that is, individuals acting without restraint will often infringe the rights of others in ways widely enough deemed unacceptable as to result in rules.

Rules cannot be enforced without threat of punishment.

Threats are not effective unless real.

When threats must be carried out, punishment will be resisted.

When punishment is resisted force becomes necessary.

These arguments, which seem (perhaps unfortunately), to describe the operational background of everyday law and order in modern societies, obviously assume that individuals and groups tend to seek to serve their own interests, and that, therefore, peace at least *usually* results from the use (or threat) of force (it obviously does not follow that *all* uses of force lead to peace). This is more an argument from empiricism than from deterministic theory.

So it appears that, if morality is actually an evolved phenomenon—a way people have worked out to serve their own interests in ways that tread on the toes of others only in acceptable fashions—then anyone who analyzes morality, who attempts to bring its cost-benefit decisions into his own and others' consciousnesses, is likely to be judged immoral both for doing it in his own mind and for trying to cause it, or risking it happening, in others' minds. He will be seen as reducing the likelihood of Frank's version of "self interest" turning into, being perceived as, or remaining as "commitment" through indoctrination and practice that produces a "suitable" social conscience. Recall Ambrose Bierce's (1911) definition of a cynic as "a blackguard whose faulty vision sees things as they are, not as they ought to be." The requirement that social motivation be nonconscious may turn out to be the most remarkable and restrictive of all moral rules, enforced by the extreme cleverness of humans at ferreting out what is in an associate's conscious mind, as opposed to his unconscious mind, and by the maintenance of heavy penalties for deliberate deception because deliberate deception

implies the greatest danger to the deceived, and repetition when opportunities arise. In other words, this situation is in many ways the most important one involving reputation. Gibbard writes as follows:

> Singer (1981) discusses systems of unpaid blood donation as working refutations of the contention that altruism can only exist among kin, within small groups, or where it pays off by encouraging reciprocal altruism. Alexander (1987) offers an explanation in terms of indirect reciprocity: the blood donor shows himself altruistic, and so "he may receive his 'payment' from the members of society who accept him in social interactions or treat him deferentially." This raises a number of questions. Why will others specially accept the unconditional altruist or defer to him? In hope of gain? An unconditional altruist would be one who showered goods irrespective of whether other accepted him or deferred to him. Blood donation must signal something other than this, but how? Then too, there are puzzles about signals in general. Why not signal the usefulness of accepting me or deferring to me, and then save myself the trouble of delivering the expected reward? And if everybody does that, why should anyone take blood donation as a sign it will pay to cultivate me? (1990:260, footnote)

Actually, my discussion was designed to show, not that a blood donor is an unconditional altruist, but that he explicitly is a conditional altruist. It is difficult to know whether Gibbard realized this or not. By my argument, blood is given only when potential, possible, likely rewards exist, whether or not they are consciously understood by anyone. The main reward is reputation, and all the benefits that high moral reputation may yield. Reputation as an altruist pays: everyone gains by the presence of beneficent people and the possibility of interacting with them, and so we tend to cultivate and reward them. In his text, Gibbard goes on to explain precisely this (p. 261): "A sense of fair dealing prompts one to cooperate, and that elicits cooperation from others. Gratitude prompts the kinds of actions that will draw more favors. Retaliation deters. These sentiments pay . . ." He then argues that judgments about fairness have to mesh among social cooperators, and that they tend to be established by discussion and interaction. He wishes to establish that our social behavior cannot be preset by the interaction of genes and parts of the environment other than our actual sociality. I expect that we all agree with him. Would that this could end the debates about determinism.

Returning to Gibbard's footnote, the reason it won't work for me just to tell you how good I am as a reciprocator, and you me, and for all of us then to act on this information, is that we are all too clever for that. We do not always mean what we

say. And we also deny that we deceive—both to others and to ourselves. If Gibbard really believed what he wrote in his footnote, then one would have to expect that he would always lose in a horse trade or when buying a used car. Horse traders and used car salesmen invariably tell us what good reciprocators they are, but we had better form our own opinions or stay home.

Horse traders and used-car salesmen also tend to be notorious as examples of unscrupulous people—people who are not moral. Why? Because we imagine that they are not only out to serve their own interests but that they do what they do from entirely conscious motivations. They know what they're doing, and they do it "on purpose." This is apparently, in part, the implication that causes Frank (1988) to substitute "self interest" for "conscious self interest."

Then do the rest of us—all of us moral folk—not know what we are doing? Is that what makes us moral? Are we moral when we serve our own interests if we actually think we are serving the interests of others? Is that all it takes? In the eyes of others, pretty much so, I would say (Alexander 1987, Table 2.5, pp. 98-99), and if that is true, it means that we had better keep it that way if we wish ourselves to be judged moral.

I have now come around to the idea, expressed at the outset, that evolution is denied (one way or another) by many who know about it because of intuitive feelings about human motivation. I believe that evolutionary biologists who attempt to explicate human behavior are ignored or maligned, in many instances, because of a widespread belief that if their analyses are correct or become public the effects will be contrary to whatever people in general believe would be beneficial. We are returned to the exclamation of the wife of the Bishop Wilberforce, upon hearing of Darwin's view of evolution, that we should pray that it not be true and if it be true pray that it not become generally known.

What is the situation, then, in which the concept—the idea, practice, or realization—of morality arises? First, social reciprocity of the direct sort must be an integral part of sociality. That may not be the case in any species except our own. What is required are indefinitely continuing interactions between intelligent beings in which each can benefit from cooperating with the other, and interactions in which defection or failing to give the share that will cause the other also to benefit are possible but will in the long run represent net losses to the defector. All of this requires a great deal of social learning. Social reciprocity depends on social learning because participants in a socially reciprocal interaction change continually and only the quick and capable learner—probably only the competent and clever scenario-builder who

knows and understands his interactants—can continue to cause the interaction actually to be reciprocal. So sometimes we know that we attempted a selfish act through a seemingly beneficent one and failed—failed, that is, to receive a return benefit fitting for the occasion. And we all know that even this kind of consciousness about kindnesses tends to be viewed with suspicion or skepticism. It borders closely enough on being immoral as to require great tact and care in its expression.

The second requirement for the concept of morality to arise and permeate social interactions is for *multiple* parties to be cooperative in the way that *two* parties can be, so that multiple cooperative associations are possible and reputation becomes crucial. To say that reputation becomes crucial means that those who become known as defectors lose—they are less able to locate and engage in the cooperativeness that benefits nondefectors. It means that there are multiple possible cooperators available for all so that changing associations can occur and the defectors—the immoral—can be shunned and left out, and as a result lose. When, and if, we are able to create a permanent situation like this among the nations of the world as well as within nations and smaller groups, then, I suggest, we will have reduced the problem of arms races and international competition to a manageable state.

Finally, I will attempt to state more precisely the paradox inherent in the nature of human intent and its relationship to consciousness. First, we relegate to the nonconscious cost-benefit decisions about particular social acts, perhaps in a fashion similar to our relegation to the nonconscious relatively simple acts such as typing or the playing of a musical instrument. This we can understand as part of a system for becoming able to respond quickly and certainly in ways that tell others we are likely to be good reciprocators—good social interactants. In other words we demonstrate that we are prepared to be beneficent without thinking about it, without a cold calculation of costs and benefits (Alexander 1987; Frank 1988): we develop a social conscience. Second, our nonconscious is evidently designed to expose to our social interactants at least some evidences of our honesty or lack of guile: thus we blush "inadvertently" in situations and fashions that also indicate to others that we are programmed to behave honestly or fairly or innocently. Perhaps this is a system for suggesting to others that we have acquired the already described tendencies to behave quickly and certainly as good reciprocators. Third, and considerably more problematic, our nonconscious also seems designed to expose at least some of its evidences of our *dishonesties*! We blush and shift our eyes and smile and do other things that reveal to others when we are lying or failing to tell something that we know. We do such things while we seem (to ourselves) to be striving as hard as possible not to. This is more difficult to explain, but

it may represent a window into our mental activities that also implies programming for honest, guileless behavior. Knowledge of the particular situations and behaviors involved in such inadvertent expressions of intent is needed to test this question; should such behaviors appear only when a strong likelihood of being discovered in a deception exists, or when the cost of being discovered is greatest (perhaps at that moment when the conscious self is otherwise about to make a costly mistake), they could indeed have an effect beneficial to the inadvertent revealer of emotions. Fourth, and most problematic, we use the conscious and nonconscious aspects of our mentality in the particular fashions I have just described, even though consciousness itself, and the building of mental scenarios, is increasingly regarded as having evolved as social machinery; that is, as having been favored because it enabled us to manipulate others in the service of our own interests by projecting alternative behaviors that interpret and manipulate social sequences as they unfold. How does it happen that we gain from removing crucial aspects of social communication from the conscious, or perhaps never actually gaining control over nonconscious social communication, especially if consciousness evolved as a tool of social communication and a predictor of the actions of others? Is it primarily a question of timing of revealing responses? I believe that when such questions are answered, human behavior can begin to make sense in terms of an inclusive-fitness-maximizing goal for the human organism, and a significant barrier to human self-understanding will be removed.

Conclusions

The evolutionary approach to human behavior has serious problems. The most obvious is that evolution is accepted by only a minuscule proportion of the world's population. More importantly, and probably contributory, even within that minuscule proportion, many academicians and others among the intelligentsia tend to wall evolutionists off like a malignant tumor. This happens partly because a minority of thoughtful people are educated in biology, and this in turn is partly because the interaction of heredity and development is so poorly understood within biology. Also involved is the convoluted nature of human efforts to self-understand, owing not only to the fact that we must use the properties we wish to understand to carry out the analysis, but also to the nature of the evolutionary (selective) history that no one wants to hear about. It is not easy for anyone to believe, from his own thoughts about his personal motivation and that of other humans, that humans are designed by natural selection to seek their own interests,

Richard D. Alexander

let alone maximize their own genetic reproduction. If natural selection is being interpreted accurately by modern biologists, then it appears to have designed human motivation in social matters as to cause its understanding to be resisted powerfully.

We lose, in analyzing such problems, if we restrict ourselves to discussing only the brighter side of human nature or pretend that the topic is cooperation *not* competition. Moral philosophers and other academicians for the most part live in pleasant worlds, with little opaque clouds that tend to admit only the delightful aspects of human intentionality floating above their heads as they move along the sidewalks of urbania between their offices and their homes. But the misery in the world is not all there because of pathologies easy to understand or proximate causes easy to remedy; nor is it all owing to those "other" kinds of people whose motivations, unlike our own, are pernicious and self-serving. Moreover, civilization and technology have created circumstances in which virtually all human striving, designed as it is to better the "current" quality of life, threatens increasingly the (more distant) future of humans, or even of life itself.

Analysts of morality must retreat from their subject far enough to examine the reasons for its convolutedness. We must know the ways in which kindness, beneficence, and good fellowship can be selfish, and we must also understand why this idea is repugnant and what if anything to do about that. Most important, to solve the problems human evolutionists have glimpsed so far, we must enlist a far greater proportion of the world's thinkers. If, as knowledgeable people increasingly suggest, massive altruism by our generations will be required to ensure the survival of later generations, then, unless we don't care, we have to know how to modify the relevant aspects of the striving we have evolved to accomplish. We have to know how to escape our history. No part of biological theory has ever legitimately implied that humans cannot employ their evolution-given traits to set and accomplish goals that are purely incidental—even contrary—to their history of natural selection. My arguments suggest that these things will happen only when evolution-minded people have overcome resistance to evolutionary analysis of behavior by explaining, much better than they have so far, the nature of human motivation and the reasons for its partial concealment and seeming convolutedness.

Acknowledgments

References

Alexander, R. D. 1969. Arthropods, 167-219. In *Animal Communication*, ed. T. A. Sebeok. Bloomington, IN: University of Indiana Press.

Alexander, R. D. 1977. Evolution, human behavior, and determinism. *Proceedings of the Biennial Meeting of the Philosophy of Science Association* (1976)2:3-21.

Alexander, R. D. 1979. *Darwinism and Human Affairs*. Seattle, WA: University of Washington Press.

Alexander, R. D. 1987. *The Biology of Moral Systems*. Hawthorne, NY: Aldine de Gruyter.

Alexander, R. D. 1989. The evolution of the human psyche, 455-513. In *The Human Revolution. Behavioural and Biological Perspectives on the Origins of Modern Humans*, ed. P. Mellars and C. Stringer. Edinburgh: University of Edinburgh Press.

Alexander, R. D. 1990. Epigenetic rules and Darwinian algorithms: The adaptive study of learning and development. *Ethology and Sociobiology* 11:1-63.

Alexander, R. D. 1991. How Did Humans Evolve? Reflections on the Uniquely Unique Species. Ann Arbor, MI: *University of Michigan Museum of Zoology Special Publication* 1:iii + 38 pp.

Alexander, R. D. In press. Social learning and kin recognition: An addendum. *Ethology and Sociobiology*.

Alexander, R. D., and G. Borgia. 1978. Group selection, altruism, and the levels of organization of life. *Annual Review of Ecology and Systematics* 9:449-74.

Axelrod, R. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Axelrod, R. 1986. An evolutionary approach to norms. *American Political Science Review* 80:1095-1111.

Barkow, J., L. Cosmides, and J. Tooby, eds. 1991. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Bierce, A. 1911. *The Devil's Dictionary*. New York: Crowell.

Boyd, R., and P. J. Richerson. 1985. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Cosmides, L. 1989. The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31:187-276.

Cosmides, L., and J. Tooby. 1988. Evolutionary psychology and the generation of culture. Part II. Case study: A computational theory of social exchange. *Ethology and Sociobiology* 10:51-97.

Daly, M., and M. Wilson. 1981. Abuse and neglect of children in evolutionary perspective, 405-16. In *Natural Selection and Social Behavior. Recent Research and New Theory*, ed. R. D. Alexander and D. W. Tinkle. New York: Chiron Press.

Daly, M., and M. Wilson. 1988. *Homicide*. Hawthorne, NY: Aldine de Gruyter.

Darwin, C. R. [1859] 1967. *On the Origin of Species.* A facsimile of the first edition with an introduction by Ernst Mayr. Cambridge, MA: Harvard University Press.

Dawkins, R. 1982. *The Extended Phenotype.* San Francisco: W. H. Freeman.

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

Dawkins, R. 1989. *The Selfish Gene.* New Edition. Oxford: Oxford University Press.

Dawkins, R., and J. R. Krebs. 1978. Animal signals: Information or manipulation? 282-309. In *Behavioural Ecology*, ed. J. R. Krebs and N. B. Davies. Oxford: Blackwell Scientific Publications.

Dobzhansky, T. 1961. Discussion, 111. In *Insect Polymorphism*, ed. J. S. Kennedy. London: Symposium of the Royal Entomological Society of London 1.

Fisher, R. A. [1930] 1958. *The Genetical Theory of Natural Selection.* 2d ed. New York: Dover.

Fletcher, D. J. C., and C. D. Michener, eds. 1987. *Kin Recognition in Animals.* New York: John Wiley and Sons.

Flinn, M., and R. D. Alexander. 1982. Culture theory: The developing synthesis from biology. *Human Ecology* 10:383-400.

Frank, R. H. 1988. *Passions within Reason: The Strategic Role of the Emotions.* New York: W. W. Norton.

Gautier, D. 1990. *Moral Dealing: Contract, Ethics, and Reason.* Ithaca, NY: Cornell University Press.

Gibbard, A. 1990. *Wise Choices, Apt Feelings. A Theory of Normative Judgment.* Cambridge, MA: Harvard University Press.

Gould, J. L. 1986. The biology of learning. *Annual Review of Psychology* 37:163-92.

Haldane, J. B. S. 1955. Population genetics. *New Biology* 18:34-51.

Hamilton, W. D. 1964. The genetical evolution of social behaviour. I, II. *Journal of Theoretical Biology* 7:1-52.

Hamilton, W. D. 1966. The moulding of senescence by natural selection. *Journal of Theoretical Biology* 12:12-45.

Hayles, N. K. 1990. *Chaos Bound: Orderly Disorder in Contemporary Literature and Science.* Ithaca, NY: Cornell University Press.

Hepper, P. G., ed. In press. *Kin Recognition.* Cambridge: Cambridge University Press.

Hinde, R. A., and J. Stevenson-Hinde, eds. 1973. *Constraints on Learning: Limitations and Predispositions.* New York: Academic Press.

Holmes, W., and P. W. Sherman. 1983. Kin recognition in animals. *American Scientist* 71:46-55.

Humphrey, N. K. 1976. The social function of intellect, 303-18. In *Growing Points in Ethology*, ed. P. P. G. Bateson and R. A. Hinde. London: Cambridge University Press.

Huxley, T. H. [1896] 1909. *Autobiography and Selected Essays.* New York: Houghton Mifflin.

Johnston, T. D. 1982. Selective costs and benefits in the evolution of learning. *Advances in the Study of Behavior* 12:65-106.

Johnston, T. D. 1985. Learning and evolution. *Scientific Progress* (Oxford) 69:443-60.

. Lloyd, J. E. 1977. Bioluminescence and communication, 164-83. In *How Animals Communicate*, ed. T. A. Sebeok. Bloomington, IN: Indiana University Press.

Lumsden, C. J., and E. O. Wilson. 1981. *Genes, Mind, and Culture: The Co-Evolutionary Process*. Cambridge, MA: Harvard University Press.

Murphy, J. G. 1982. *Evolution, Morality, and the Meaning of Life*. Totowa, NJ: Rowman and Littlefield.

Otte, D. 1974. Effects and functions in the evolution of signaling systems. *Annual Review of Ecology and Systematics* 5:385-417.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Rodseth, L. 1990. Prestige, hunger, and love: Plumbing the psychology of sexual meanings. *Michigan Discussions in Anthropology* 9:1-21.

Sahlins, M. 1965. On the sociology of primitive exchange, 139-236. In *The Relevance of Models for Social Anthropology*, ed. M. Banton. London: Tavistock.

Simon, H. A. 1990. A mechanism for social selection and successful altruism. *Science* 250:1665-68.

Skinner, B. F. 1965. *Science and Human Behavior*. New York: The Free Press.

Symons, D. 1987. If we're all Darwinians, what's the fuss about? 121-46. In *Sociobiology and Psychology*, ed. C. Crawford, M. Smith, and D. Krebs. Hillsdale, NJ: Erlbaum Publishers.

Thorpe, W. H. 1950. The modern concept of instinctive behaviour. *Bulletin of Animal Behaviour* 1(7):2-12.

Thorpe, W. H. 1951. The definition of terms used in animal behaviour studies. *Bulletin of Animal Behaviour* 1(9):34-40.

Tierney, A. J. 1986. The evolution of learned and innate behavior: Contributions from genetics and neurobiology to a theory of behavioral evolution. *Animal Learning and Behavior* 14:339-48.

Tooby, J., and L. Cosmides. 1988. Evolutionary psychology and the generation of culture. Part I. Theoretical Considerations. *Ethology and Sociobiology* 10:529-49.

Trivers, R. L. 1971. The evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35-57.

Waddington, C. H. 1956. *Principles of Embryology*. New York: Macmillan.

Waldman, B., P. C. Frumhoff, and P. W. Sherman. 1988. Problems of kin recognition. *Trends in Ecology and Evolution* 3:8-13.

West-Eberhard, M. J. 1987. Flexible strategy and social evolution, 35-51. In *Animal Societies: Theories and Facts*, ed. Y. Ito, J. L. Brown, and J. Kikkawa. Tokyo: Japan Scientific Society Press 1.

Williams, G. C. 1957. Pleiotropy, natural selection, and the evolution of senescence. *Evolution* 11:398-411.

Richard D. Alexander

Williams, G. C. 1966. *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press.

Williams, G. C. 1989. A sociobiological expansion of *Evolution and Ethics,* 179-214. In *Evolution and Ethics*, ed. J. Paradis and G. C. Williams. Princeton, NJ: Princeton University Press.

Williams, G. C., and D. C. Williams. 1957. Natural selection of individually harmful social adaptations among sibs with special reference to social insects. *Evolution* 11:32-39.