

**Citation:** Lahti, D. C. 2013. Twelve (more) things about the evolution of morality that make people nauseous. In K. Summers and B. Crespi (eds.), *Human Social Evolution: The Foundational Works of Richard D. Alexander*. New York: Oxford University Press, pp. 307-325.

## Twelve (more) things about the evolution of morality that make people nauseous

David C. Lahti

“Over there they believe society is based on lies, but here we believe in justice.” That is essentially what I was told by an experienced graduate student at the University of Michigan when I mentioned that I was going to visit Dick Alexander. When I arrived in the orthopteran range at the Museum of Zoology, eight or nine backs were hunched over a new *Nature* paper looking for a fatal error (Keller & Ross 1998). It seems that a “green beard”—a previously hypothetical trait that is claimed to indicate relatedness accurately and that individuals can recognize in each other and use as a signal for preferential treatment—had been discovered in the red fire ant *Solenopsis invicta*. Perhaps the experimental odor transfer control hadn’t been done perfectly. Perhaps they used the term “outlaw gene” (Alexander & Borgia 1978) too loosely. But all the elements were there! Queen ants that reproduce have a genotype *Bb*. Queens with a *BB* genotype attempting to reproduce get killed, primarily by those with the *Bb* genotype, who can tell the alleles apart via cuticular hydrocarbons. Fine, it’s probably a real green beard in nature. But theorists have warned that such a system is vulnerable to cheating. A *BB* ant might eventually masquerade with the *Bb* signal, if it is possible to separate those two effects of the gene. Deceit would be adaptive for *BB* ants. Later I hesitantly raised the objection that any deceit, or even concealment for that matter, might be hard to maintain long term in the face of others’ interests in uncovering the truth. “Does your wife know when she is ovulating?” retorted one graduate student. No, I admitted. Of course she can learn to do so, but only because of modern medical and physiological advancements that have revealed the very existence of ovulation. And yet how important ovulation is to reproduction! And how flamboyantly some other female apes advertize it! So why do we conceal it? Very interesting. Then I read the so-called “Uniquely unique” paper (Alexander 1990), which had been published as an unprecedented Special Publication of the Museum of Zoology because it was deemed too controversial for the ordinary Occasional Publications (*You have got to be kidding me*, I thought—but I suppose I’d rather a paper of mine be considered Special rather than merely Occasional). I was shortly convinced that hiding the truth, not only from mates but even from oneself, was adaptive for female humans in the case of ovulation, as it made males more likely to pair-bond and provide biparental care, features so important in our socially competitive environment. Adaptive deceit in ants, adaptive evasion in humans: *We’re already getting close to a society built on lies and I’ve only been here one day. No wonder the ecologists for social welfare across the street warned me. But it’s too late now.*

Why do we see resistance to understanding ourselves evolutionarily, even among those (like those ecologists) who broadly understand and accept evolution? Alexander, whose synthetic evolutionary approach to human culture is the most extensive, careful, and consistent around, has returned to this question repeatedly, especially in recent years (Alexander 1978; Alexander 1988; Alexander 2005; Alexander 2009). I’d like to focus this question further here and apply it specifically to our understanding of the origins of morality, especially the understanding that has been promoted by

Alexander's own work. The paper "Biology and the moral paradoxes" (this volume, and Alexander 1982) is a succinct and effective summary of the aspects of Alexander's thought that can lead us into the resulting controversy and resistance.

One fascinating and frustrating phenomenon is that those who study morality professionally are among the least likely to understand or accept our best theory of the evolution of morality. I would argue that this rejection is not primarily because of anything they know about morality that evolutionary biologists do not (although that is certainly the case, and a problem that should be redressed). A more pervasive and general reason is simply that most philosophers and social scientists still practice their arts as though humans are not a product of evolution. This is astonishing and becomes more so with each passing decade. However, aside from their attitude that evolution in general is irrelevant, their specific loathing and dismissal of the emerging picture of the evolution of morality needs special explanation; and I argue that *this* attitude is not at all surprising and is likely to continue for a long time. Even many biologists who do not think much about human evolution have a similarly negative response to evolutionary moral theory, almost by default. I think the main reason for the nausea is that our best theory of the evolution of morality looks as though it were concocted precisely for the purpose of slapping moralists in the face, making fun of goodness, and dragging through the mud everything pure and decent in the world.

Before we get to this, however, there are several arguments the professional morality folks use to stop evolutionary analysis before it even starts. Most of these have already been discussed to death and I will not deal with them again here: a group of doubts about the explanatory efficacy of science in general or evolutionary science in particular, including concerns about the nature of history, social and other influences that compromise objectivity, and the cute postmodern idea that science might simply be one of many possible narratives; another family of concerns encompasses metaphysical or religious (including political and social theoretical) taboos about "human nature", including a preference for absolute freedom of the will, worries about genetic determinism, throwback Marxist hopes, or just a general distaste for inherited influences on the human mind. These venerable arguments could be parsed into a dozen perceived problems with an evolutionary approach to morality. I will leave those aside here in favor of another dozen. Actually, the first two remain in this general show-stopper category, often surfacing as peremptory blurts. I will consider them only briefly: that evolution may well happen but it is irrelevant to philosophy, and that a lack of consensus is an excuse not to think about the issue.

Philosophers frequently sniff at **(1) the outrageous suggestion that evolutionary biology has anything at all to say about morality**. Such a radical separation of science and philosophy would have been alien to nearly all the great philosophers of old, including Aristotle, Descartes, Hume, and Kant, who were inspired by the latest empirical knowledge and stretched its implications as far as they could into a springboard from which to leap into realms that science had not yet fathomed. A few philosophers still retain this idea of science as handmaiden to philosophy. How different Wittgenstein, who was so content with a narrow conception of philosophy that he could write (and I am so sick of seeing this quoted) that "Darwin's theory has no more to do with philosophy than any other hypothesis in natural science" (Wittgenstein 1921). I respect Wittgenstein's experiment in philosophical minimalism, in the

spirit of the beginning of Descartes' *Discourse on Method* (1637). But should philosophy really be nothing but conceptual clarification? No. Philosophy, φιλοσοφία, is literally the love of wisdom, it is the search for answers to the big questions. And it has a responsibility to take all human observation into account in order to do this. Your concepts themselves, philosophers, depend on your psychology, that is to say your minds, which are a product of history and have evolved genetically and culturally according to particular rules for particular reasons in particular environments. Your concepts would be different if any of those things were different. To practice philosophy as if conceptual clarification can be done in ignorance of human natural history is to assume an idiosyncratic human natural history of your own, one that probably dates back to Hume or even to Plato, and represents the best of what they knew about humans. To engage in metaphysics without understanding human evolution is to assume a false pre-evolutionary anthropology. To engage in moral philosophy without understanding human evolution is to assume a false pre-evolutionary psychology and sociology.

Alexander, building on the progress of other evolutionary biologists such as Darwin, Williams, Hamilton, and Trivers, has birthed an expansive and vibrant philosophy, representing the first several steps in a re-evaluation and recasting of the world's big questions in the light of evolution, including those relating to morality (e.g., Alexander 1979; 1982; 1987; 1989). Of course Alexander, like many scientists, considers it a slight for his work to be referred to as *philosophy*, which everybody knows is a code-word for a pompous and meticulous web of utter speculation. This is not philosophy at its best, however. Philosophy is an art of precise and internally consistent hypothesis generation. And the more there is to be explained, the more good philosophizing is necessary in order to develop a hypothesis that is broad and detailed enough to explain it. Philosophizing about something as formidable as human nature must amount to the generation of a large ball of interacting hypotheses, all of which are consistent with each other and explain all observations to date.

Another case rather shamelessly proffered by professional thinkers about morality is that **(2) the difference of opinion among evolutionary biologists** as to the explanatory power of their field with respect to morality gives the rest of the world an excuse to let the matter sit. Certainly there are many theories out there, and everyone has a twist: cognitive scientists, psychologists, a diverse array of anthropologists, primatologists, economists, journalists, and yes zoologists. Some treatments are better than others, most are rhetorically overblown and oversimplified, and (to be frank) few of them deserve to have seen the light of day (Lahti 2003). If the students of evolution can't even get it straight among themselves, why should philosophers and social scientists be concerned? Of course, this logic is especially dangerous for philosophers to use because there is not a single important matter in the world on which there is a reasonable consensus among their own ranks. The fact is, although science does proceed socially by consensus, no philosopher nor scientist can get any sort of answer to any big question by depending on, much less waiting for, consensus. The bigger the question, the more difficult wrapping one's head around it becomes, the more axes are out there to grind, the more diverse is the field of ideas, and so the poorer the consensus. Humans will always fret and explore most intensely about issues that have to do with humans—that's almost a truism. Rather than letting the matter sit, we must compare the extant hypotheses, winnowing them for consistency, explanatory power, and their ability to pass tests having the potential to falsify them.

Provided that evolutionary analysis of morality is actually allowed to begin, any serious work might very well meet with a turned-up nose regardless of what the final picture looks like, because of a three-pronged arsenal of arguments, almost like an immune defense. First we have the lingering effects of a prior infection: **(3) guilt by association with “social Darwinism”**. In the early days, some theorists attempted to derive morals from the evolutionary process, some of which claimed some humans to be better than others and justified power disparities and social injustices. Such detestable ideas had the effect of virtually inoculating the academy against susceptibility to any future outbreaks of evolutionary thinking about morality. Even today, the mere mention of the term “social Darwinism” exerts great rhetorical power, whether it is used appropriately or not. If for some reason that prophylactic doesn’t do the trick, the second line of defense against evolutionary ethics is to take two fallacies and call a philosopher in the morning. The two fallacies in question do highlight actual errors of thought when used properly, but with regard to the connection between evolution and morality they can be portrayed as having mighty accusatory powers far beyond their logical reach. One is **(4) the genetic fallacy**, where the origin of an idea is used as an argument about its validity or truth. All you have to do (some think) is accuse someone of this, and they will be prevented from considering the origin of morality to have any relevance for... morality. Wielding **(5) the naturalistic fallacy** can be just as vital. This fallacy is committed whenever (depending on its formulation) someone claims that moral values are implied by, are defined as, are really the same thing as, or somehow arise out of, facts of the sort that natural science can countenance. Some take this fallacy to absolutely debilitate any attempt to make morality natural, or rooted in things that are accessible to natural science, which presumably includes anything that has evolved. With this arsenal in place, the professional students of morality might not have needed much else in order to protect their subject from evolution. (Someone should name a fallacy for the reckless wielding of fallacies.) However, strange as it may seem, these defenses against evolutionary biology’s intrusion were probably not even necessary: evolutionary biology became its own worst possible publicist.

To start with, although evolutionary biology does not preach a morality, it describes the human condition in the context of that of all life, where traits persist and spread insofar as they benefit their individual bearers. Thus, any consistent explanation of human action from evolutionary biology will look like **(6) egoism**, which is often considered “knockabout philosophy” as my philosophy advisor used to use the term—a view you toss around to whet the critical abilities of young thinkers, but eventually discard for more serious contenders. Still, there are always plenty of “enlightened” egoists around, people who believe that in some way what is good is so because (or consists in the fact that) it is good for us. The horrifying part is the particular meaning of “good for us” that evolution brings to the table. This is where evolutionary biology irreparably ruins its image. What we humans consider good, we came to consider such because... it is good for... **(7) reproductive success**. This means, for starters, *sex*. After thousands of years of morality being a vanguard against... well, immorality if you know what I mean, we are asked to believe that the two have always been in bed with each other! The idea that moral goodness or rightness has any sort of basis at all in making babies—no, worse, in making *more babies than other people do*—is probably the most odious thing any thinker could suggest about the nature of morality. Or, if there were a more repulsive connection, it would have to be the other thing we protect our children from in the name of goodness: **(8) violence**, and especially *war*. And lo and

behold, this becomes in the evolutionary picture the sine qua non of morality! All of nature struggles to reproduce, but only humans have morality. Why? Because we have lived in social groups that have competed so fiercely with each other that it led to arms races, not only of literal weapons but also, and more crucially, of minds. We outsmarted other hostile forces to become each other's worst natural enemy. And the group competition was so unrelenting that the social group had to be unified or else disintegrate. Any individual's successful reproduction would depend on the persistence of the group, necessitating cooperation with other group members, and standards for such cooperation, including to some extent the sacrifice of immediate individual interests for the greater good. Thus, paradoxically, war led to goodness, and without war we would never have evolved the intellectual ability to decry war, nor to hope for something better. Moreover, the mechanism by which this happened was a "selfish" maximization of individual reproductive success.

What should be our response to repugnance at this history? In my opinion, our response should be to encourage quality science education so that the repugnance declines with time. As morally ambiguous as this proposed history is, it is our most powerful explanatory framework for human nature. Those who understand natural selection and appreciate its implications, and who are aware of the nature of social behavior in other species, are not likely to encounter a tremendous barrier to realizing or accepting this picture. And at this point, despite evolution-inspired nihilistic popularizations, those who respect morality and the idea of goodness might still rescue them in dignity from the rubble of their history. After all, egoism in an evolutionary sense includes concern for one's kin and group, reproductive success for humans includes not only sex but pair-bonding and parental care, and fighting only sowed the seeds for morality insofar as it protected one's social group and became the wellspring of cooperation. More generally, many human traits derive from precursors that were humbler in some (anthropocentric) sense. Just as many of us have risen above a Wilberforcean indignance at the idea of having evolved from animals that don't wear clothes or speak very well, maybe we can stave off moral discomfort at the centrality of selfishness, sex, and violence in our moral evolution.

\*\*\*\*\*

This is not the end of the story, however. The foregoing is prudish stuff compared to the next level at which moral ambiguity enters the evolutionary picture, the full extent of which many philosophers and social scientists are not even aware, so poor is the diffusion of ideas across disciplinary borders. This is the level of the individual human psyche, and of the motivations or intentions that guide our actions. For morality to have fulfilled its ancient and continuing evolutionary function (thus explaining its persistence in the repertoire of the human species) we can expect **(9) a correlation between the interests of the genes and the goals driving human action**. To the extent that reproductive success has been the currency of natural selection, it will accordingly have evolved to be a predominant and universal human aim. And to the extent that group competition has been the engine of cooperation, corresponding desires and motivations will have evolved in our psyche as well. Moreover, any exceptions or supplements to this picture are not expected to be universal or particularly lasting. Views and people going against this trend will be statistical residuals in the long view. What this means for human psychology and personality is morally complex: we are expected to act primarily to benefit ourselves individually, as well as our mates and offspring, and (because of indirect fitness benefits) our

nondescendant kin. We will tend to value members of our own various social groups over nonmembers, and we will tend to withhold beneficence and be less caring the more distantly related people are to us. Some of these tendencies and the implications from them we might consider morally praiseworthy, others acceptable, and many we would consider immoral. There is no consistent explanation for the evolution of moral behavior—rather, there is the explanation for the evolution of *human* behavior, including the moral and immoral mixed up together.

To such standard material from the Alexandrine picture of human behavior we must add still more troubling consequences from the interaction between group competition and indirect reciprocity. Indirect reciprocity, the mechanism by which rewards and punishments can return to someone through any individual in society or even from society as a whole, is the most pervasive and explanatory element in the evolutionary analysis of morality (see Sigmund, this volume). While still yielding falsifiable predictions, indirect reciprocity renders unproblematic a majority of the scenarios that are typically offered to challenge to the evolutionary picture, from blood donation and philanthropy to monasticism and celibacy. Considering the benefit of group unity, indirect reciprocity becomes a vehicle for widespread social pressure to contribute service to the group. Not surprisingly, much of what we consider moral or good tends to be consistent with this rubric of service to the group. Insofar as group cohesion preserves the group, it benefits the reproductive success of its members; consequently these members will tend to benefit by serving the group, encouraging this group service in others, and socially enforcing their commitment. However, more broadly, each individual will also gain in reproductive success within the group by seeking benefits for oneself at the expense of others in the group. The result of this analysis is that each individual should encourage in others a greater level of commitment to group service than one's own optimal level of commitment. Moreover, by the mechanism of indirect reciprocity it is less the actual good behavior or intentions that lead to reproductive success, and more the *perception* by others that one is good. Thus we are likely adapted to inflate the impression we lend to others of our goodness. The evolutionary prediction in short is that **(10) hypocrisy is inherent in human nature and an essential part of adaptive moralizing**. Many adaptive strategies can be drawn from this prediction, including ways to discriminate in our dispensing of benefits, to play upon the altruism of others, and to anticipate and manipulate others' impressions of us for our own gain. The discussion of morality has turned into a discussion of how we get around morality, and the evolution of goodness has turned into the evolution of the best tool in the con man's kit. Thoroughgoing goodness doesn't exist except in the sucker, and so those of us in the know have every reason to exchange evil grins and high fives (or perhaps more prudently, discreet winks) when we encounter idealistic people who really appear to have some respect for the concept.

Actually this is not quite right. According to the emerging evolutionary picture of morality, the main reason we do not see so many evil grins and high fives, or even winks, in response to others' moralizing or greater group service than our own, is more than just prudence. It is possible that some or even most of us do not generally stifle obvious feelings. Alexander's more likely alternative is **(11) self-deception**. We are not perfect at hiding our conscious motives, and others are evolved to detect them. Therefore the mechanisms just described work better when we are just as clueless as our audience is, to the reasons why we have come to encourage moral adherence and laud heroism. (Hence, Alexander adds,

our resistance to accepting those mechanisms (Alexander 1987)). We tend to be conscious of things when being conscious of them is advantageous, and not otherwise. (Note that modern psychology has it backwards when it concentrates on why we sublimate or render certain events or motives unconscious, when this is in fact the primitive and much more widespread state of affairs in nature. The real question is why our ancestors became conscious of certain things.) This incomplete and even false knowledge about ourselves allows most of us to have the comfortable sensation that we are unitary and consistent in the causes of our actions. Many lines of research have since converged on this idea that we humans are inherently inconsistent and either cognitively fragmented or self-deceived. The idea has an array of explanations in the literature besides the facilitation of moralizing: is it also because we have old and young parts of our brains, or two lateral halves (Haidt 2006)? Or is it because we have some genes inherited from dad and some from mom and these can predispose us differently (Burt & Trivers 2006)? Or is it because our brain is comprised of modular neural networks that evolved for different and often contradictory functions (Kurzban 2010)? Whatever factors besides selection for moral manipulation contribute to our “impurity of heart” (Kierkegaard 1846), the upshot for our psyche is that we are far from the internally consistent and honest wills that we generally consider ourselves to be. Worse, in a sense our hypocrisy probably runs even deeper than a temptation or felt preference—it may run so deep that it is cognitively inaccessible to us, while its guidance of our action is nevertheless successfully operating in our day to day lives.

With this, even many who respect the role of evolution in human psychology and behavior have had enough. The paradox that egoism lies at the base of altruism, that self-regard somehow subsumes other-regard, is morally ambiguous enough at the level of evolutionary mechanism. For it to threaten much more seriously to undermine our integrity at the level of individual thought and action is just too much. Altruism is reduced to what we are self-deceived to believe we are being when we are actually—perhaps unconsciously—being egoistic. Any view of human thought and action that can cut through to causes beneath conscious motivation has teeth so long that an audience will be quick to conclude that this is a wolf and not grandma talking to them. The visceral responses to such a view can be potent and surprising. I know one reader of this theory who claims to have thrown *The Biology of Moral Systems* (Alexander 1987) across the room at this point. I myself admit to having written “PALTRY” in huge letters across a page of an article when I first saw this view (Alexander 1992), by which I meant that it destroyed morality and made it paltry. One respected speaker, whose lecture was entirely on the subject of Alexander’s theory of self-deception, shook his head in disgust and said from the podium, “If you can’t see what is wrong with that, I have nothing more to say to you”. A mere description of the theory once prompted an otherwise kindly ex-president of a prestigious Cambridge college to shout “NO!” and pound his fist on the table of a fancy and quiet restaurant, shaking glasses and turning heads. Surely, if one accuses humanity of wholesale lies and posturing precisely in place of everything heroic and generous, one should expect an uneasy reception. Goodness is supposed to be a heavenly ideal, something to which we can aspire with all of our being, and its behavioral counterpart rightness is born of love, the best and purest thing this world has to offer. Yet our most explanatory evolutionary theory insists these wonderful things to have arisen out of social conflict and competition, and to be saturated with manipulation, self-deception, and the danger of being found out.

Is this the extent of the better angels of our nature? Where is the goodness that moral philosophies and religions advertize and that we are apparently evolved to display and encourage in others? Where does a deep-seated moral integrity fit in this theory? The answer appears to be that it fits nowhere, or at least does not demand a great deal of scientific explanation because we do not expect such a phenomenon to be very common. We are not *Homo bonus* or *beneficiens*, but *Homo sapiens*: thinking, clever, calculating, option-weighting humans. Evolutionary biology explains the existence of the moral ideals and also explains why our motivations and behaviors do not match those ideals. In contrast, perhaps most who discuss the biology of prosocial behavior leave us with the idea that niceness is the norm and the rare Machiavellians are the exceptions to explain and avoid (e.g., de Waal 1996; Oakley 2007; Hrdy 2009; Churchland 2011). Humans do get along remarkably well compared to a chaotic “state of nature” (Hobbes 1651), and so the nicer perspectives do have explanatory power. The less savory parts of the story should not be ignored, however. If Alexander and others (Batson et al. 1999; Trivers 2000) are correct, in terms of our adherence to the ideals of the moral point of view the goodies are the exception, if they exist at all. If there is any consistent, uncalculating, and deep devotion to the moral life in actual humans, its manifestations are odd points off the trendline, exceptions that are smoothed away by the averaging of statistical tendency. Perhaps the most nauseating thing about the evolutionary moral theory for the contemporary moralist is that when one steps away from the whole picture, we can’t help but notice that despite our best hopes and cozy thoughts, **(12) we are not fundamentally good.**

\*\*\*\*\*

The biggest and least controversial shortcoming of the evolutionary account of morality is that an explanation of morality leaves open the question, *What we are to do?*—the primary moral question of the ages and (if ancient literature is an indication) a main starting point for our curiosity about ourselves as humans. Evolutionary biology leaves this question not only unanswered but in a way less sensible than when we started, because its analysis has muddied the moral waters. In fact, one who understands the evolutionary account can be forgiven for suspecting that there may actually be no answer to this question in the deep sense in which it has historically been asked; that after an iconoclastic evolutionary analysis the things most vital to be said about morality and life-living have at last been said. We are left with ethical nihilism, and so although we still need to make practical decisions, to get too excited or evangelistic about which path we take would seem forced.

If this is the case, however, why do the ends of books and papers on the subject of the evolution of morality remain so strongly moralistic? Take the end of “Biology and the moral paradoxes”, for instance: what is the impetus for the “goal of diminishing human problems through improving self-knowledge”? And on what basis do we consider “intergroup conflict” something that “we must supercede”? Why, given the evolutionary function of moralizing, should we give credence to claims that “we seek... world peace and world law” (all from Alexander 1982)? Moralizing tendrils twine through Alexander’s writings, clinging to the very moral theory that has just revealed the sordid function of moralizing. In fact, Alexander’s works deal in increasing detail with the future of religion and hope for humanity (Alexander 2009; 2012; this volume). Values poke through the evolutionary analysis, particularly at the beginnings and ends of papers, like bold weeds encroaching on a carefully tended



garden: End mass violence and hatred of outgroups! Promote a ladder of affluence to enfranchise all!  
Demand honesty in our leaders!

What is going on in these statements and encouragements, in light of the evolutionary background? Trying to answer this question shows the difficulty of applying or testing the hypocrisy and self-deception hypotheses in any particular case. The morals could simply be the theory at work. They could benefit the author either by manipulating us into being more moral than the author so that he benefits disproportionately from our group service, or else from our lauding of his efforts. This is harsh, though exactly what the theory would claim. I do offer this possibility tongue-in-cheek, though, since such a strategy would seem comically inauspicious right on the heels of sensitizing us to such ploys. Alternatively, the morals promoted might approximate strategies that are actually reproductively advantageous for an individual in our current environment. In this case the author might be behaving altruistically by giving us a heads up, contrary to theoretical predictions—although more likely our collective appreciation of him far outweighs his cost in letting the cat out of the bag, such that our advantage and his own end up being in line with each other. If all of this is too cynical, a brighter alternative is that there is wiggle room in the system, such that goodness is not entirely something cast out as a carrot to lure donkeys into doing work for us; and that striving after an ideal other than individual reproductive success is actually something that some people, even after understanding the evolutionary account, are still willing to countenance, at least to a limited extent. But lest we be too proud, striving after such an ideal for its own sake might itself generate a pretty nice reputation, which points us again towards the possibility of hypocrisy and self-deception. Thus our interpretation is plagued by an endless cyclical regress between our behavior being captive to the evolutionary moral theory and rising above it. The solution to this regress may remain forever elusive. We have no comprehensive algorithm by which to divine the impact on reproductive success of every theory, belief, and action; and we have no intrusive psychoanalysis by which we can discern an agent's motivation underlying these same events. This is not a logical or conceptual problem with Alexander's view, as much as an inherent problem (if the view is correct) in the project of humans studying themselves.

However we might choose to interpret moralizing today, one sure fact is that morals are remarkably resilient to analysis—they survive terrible beatings. Apparently we can't kill the concept of goodness even when we riddle its evolution with scandal. The concept's very nature is to remain an unassailable ideal, regardless of its origins, history, or practice, and whatever the content we ascribe to it (Murdoch 1970). We may not live the ideal, but we can't fault it for that. Thus we can talk about the corruption inherent in moralizing and then turn right around and moralize with a straight face. Likewise we can bite the hand that fed us, using our cooperative value system that was forged in group competition to bash group competition.

There are good reasons why morals can survive the harsh treatment of an evolutionary analysis. Some of these are very basic—we are moral animals, after all, and regardless of how much we morally wander, it's not clear that most of us are capable of living and thinking in line with nihilism or even according to an ethic we invent from scratch, with all due respect (i.e., very little) to postmodernists and social constructionists. Here I will mention three other kinds of reasons—that is, rational reasons, or rationalizations—why morals don't completely lose face in light of evolutionary analysis; in other words,

three reasons why the last four nauseating elements presented above do not do morality as much damage as they may seem to threaten, and as many people (such as those giving the visceral responses above) have feared they would. In brief they are (1) that there is indeed a place for the partially and even fully honest among alternative moral strategies, (2) that self-deception means precisely that the “hypocritical” might not be hypocritical, and (3) that the whole point of an ideal is that no amount of violation necessarily destroys it.

First, I suggest that the few adaptive strategies so far discussed in the area of the evolution of moral psychology need additional development. In particular, the hypocrisy and self-deception model works as far as it goes but is too simplistic by itself. Humans operate with varying personalities and in diverse environments and do not always employ a uniform strategy for reproductive success, especially when a particular strategy is risky and prone to backfire. In this area, evolutionary theorists about morality should learn from salamanders and fish. In several species, males can opt into sexual selection and compete for mates if they have the wherewithal, or else (if they do not) they can masquerade as females and gain matings by this alternative route (Gross 1996; Brockmann 2001). If these slippery creatures can opt for different strategies based on various individual and environmental variables, all functioning in the same general pursuit of individual reproductive success, one might suspect that humans can do the same in the area of cooperation. A strategy of self-deception and aggrandizing one’s moral fiber, together with the attempt to encourage more moral commitment in others than one’s own level, comprises but a portion of a range of activities that can be adaptive in a social context. It almost goes without saying that with our abilities of perception and discrimination, behaving cooperatively in actuality, especially in the low cost situations we face every day, can be a more viable alternative than attempting to deceive or manipulate others, especially considering the house of cards that reputation can be, where a lifetime of cooperation can be undone with a few or even one detrimental act (Alexander 1987). Surely not everyone has equal powers to dissemble, whether just to others or to oneself as well. Possibly emetics #10 and #11 above are better seen as two common (and complementary) tools of the human trade, than as necessarily universal and incessant practices. Of course, actually testing for the role and importance of these psychological or subpsychological tendencies would be more helpful than the speculation I am doing here.

Second, fish and salamanders aren’t the only things that are slippery: so are words. When we use terms like “selfish”, “altruism”, “hypocrisy”, and “self-deception” in an evolutionary context, we do not mean precisely the same thing that we mean by them in social usage, where these words have their ancestral currency. In typical uses these words imply a key role for intentionality. We would never denigrate as “selfish” a person whose actions merely tended to benefit oneself, if the person’s intention was to benefit someone else or to accomplish something equally unrelated to self-benefit. Following the growth of evolutionary thought, however, we tend to repurpose existing terms for all the new conceptual spaces that have suddenly opened up. We now have various proximate (mechanistic, developmental) and ultimate (functional, historical) levels on which the identical terms can be used, and the uses at the different levels are not always consistent, nor do they always imply one another. For instance, we use “selfish” of genes as a heuristic tool, despite the fact that genes cannot possibly have the intentionality that is typically implied by that term. We are to understand that the word is being

used in a role analogous to its typical one, despite some discontinuities, like when we say that skies are “threatening” or that a rattling screw “wants” to come loose from a machine. Therefore when we read “hypocrisy”, and yet admit that this phenomenon might not be cognitively accessible to us because we are self-deceived, what we are really saying is that we might not really be hypocritical at all, but we might be doing something that is analogous to hypocrisy. Real hypocrisy is presenting an appearance of a virtuous intention that one does not really have. Hypocrisy in the evolutionary theory of morality has a range of possibilities, including this real hypocrisy as well as alternatives that do not actually qualify as such. One example is a situation where one’s intention is entirely virtuous, but where that virtuous intentionality is based partly on a series of alleles that evolved by natural selection because that psychological state was beneficial to its bearer. Another is a situation where (at least some of) one’s conscious motivations are virtuous, but they are minor in their effect on action compared to certain unconscious motivations that are self-interested and (in this case) are motivating the very same action. This situation too would not satisfy an ordinary definition of hypocrisy because, in this case again, one does actually have virtuous intentions. These arguments apply equally well to “self-deception” and “altruism” (Lahti 2003): these terms in the mouth of the evolutionary biologist are not really what they seem to be. Emetics 10 and 11 are not quite as nauseating when one takes them with this grain of salt.

Actually, the evolutionary sense of hypocrisy and self-deception leads to what I would consider to be one of the striking paradoxes in the evolution of morality: that “hypocrisy” and “self-deception”, in their evolutionary senses, instead of making us worse, make us better—in fact they are our saving grace. Their importance in the theory arises because of the possibility for inconsistency within but especially between the levels on which human behavior is controlled, the level most relevant to morality being that of intentionality. Real hypocrisy occurs when our intentions are inconsistent with how we represent them to others; and nearly all of us would agree that this kind of inconsistency is generally (a bit of moralizing now) a bad thing. But the goal of human lifetimes, insofar as we are an evolved organism, is reproductive success. In light of that, if it were not for a similar kind of inconsistency between levels of behavioral control, we would have exclusively selfish intentions. In other words, if the evolutionary function of a behavior were always manifest in our intentions, we would only be able to help others by feeling or thinking selfishly, only care by faking it. Enter the ridiculous comment that “No hint of genuine charity ameliorates our vision of society, once sentimentalism has been laid aside... Scratch an ‘altruist’ and watch a ‘hypocrite’ bleed” (Ghiselin 1974: 247). We can see the germ of truth in this, but it does not reflect how humans actually work. Precisely because of inconsistency between levels of control, a mother can feed her infant with selfless love, “genuine charity”, despite the fact that the function of her behavior is to benefit her own genetic lineage (this is considered hypocrisy in the evolutionary analysis). And from maternal love, arguably, all love has evolved. Love as devotion to another is only possible insofar as its psychological workings are shielded from the adaptive basis of loving behavior; this is what evolutionary biologists call self-deception.

Still, we must admit that our motives are impure. We cannot in good conscience rescue good conscience solely by scapegoating some unintentional level of behavioral control. The selfish function of behavior so regularly informs our conscious motivations that it is likely to bleed through even when we have noble self-impressions to the contrary. As emetic #9 above suggests, the most straightforward way

for a genome to achieve reproductive success is simply to produce desires and motivations that aim immediately at survival and reproduction. Accordingly, few would disagree that desires and motivations that relate simply and closely to survival or reproduction are among the strongest in human experience. However, our sociality often calls for a subtler route to reproductive success (for instance, via enhancing group stability or our own reputation (Lahti & Weinstein 2005)), in which case those basic desires and motivations may need reining or contravention. Even that loving mother holding an infant knows well the mixed feelings, and consequently the need for self-control, associated with giving so much of herself to one offspring. Some moralists and moral philosophies demand purity of motive or intention in order for a state of mind or a behavior to be considered good, but this situation appears to be rarely if ever actualized. That we humans are a bundle of often inconsistent motives is a lesson delivered not only by contemporary evolutionary psychologists (e.g. Kurzban 2010), but frequently in the history of wise counsel on the subject. The major religions and philosophies speak generally with one voice on this matter, thus agreeing with the last and (to some) most frustrating of the twelve points above, that we are not as good as we prefer to think. If we object to this point despite the history of introspection and now despite evolutionary prediction, we are not necessarily fighting for human dignity. In fact we may be revealing a defensiveness and unrealistically high self-image. In this the demand for purity of motive resembles the demand for free will among similarly well-meaning defenders of humanity. Most of the argument about free will stresses the contrast between it and some sense of determinism. But even if we grant a sort of agency to humans, is there any scientifically informed person today who is really prepared to defend utterly *free* will? Agency without influence? We know beyond a doubt that we have influences, both culturally acquired and inherited, and these influences sway our decisions. To demand that there is some faculty that is somehow immune to these factors is unreasonable. Any conception of the will that is worth considering will have to take this diversity of interests seriously into account. In the same way, defending a moral psychology where our motives and intentions operate in a sort of cleanroom isolated from all compromising or conflictual influence is untenable and reactionary. The spirit of the evolutionary account, and very likely literally true, is that even our noblest attitudes and actions are moved not only by worthy but also by morally neutral and even base influences. When I dissect my own emotional objection to the suggestion that posturing and manipulation are an inherent part of the human condition, I wonder how much of my objection is honest doubt as to its truth, and how much is my wish that I and *a fortiori* others not be that way. I also admit an uneasy sense, as predicted by the theory, that I don't want open discussion of this tendency to lead to the spread of an opinion that these strategies are acceptable.

This brings us to the third reason why morals can survive the evolutionary account: because our own moral rectitude is not, should not be, and never really was a condition for a functioning morality. Knowing that we are bad (to whatever extent we are) does not wreck the ideals we hold up, or prevent us from holding them up, any more than denying something makes it false. Nevertheless, the fact that tendencies we would consider good evolved hopelessly tangled up with those we would consider bad does complicate our job as moral beings. It means we are stuck assigning moral values to attitudes and actions regardless of their original evolutionary function. We may praise certain things, such as sacrifice for our social group, and despise others, such as racism, regardless of the fact that these two tendencies became adaptive in the context of precisely the same function: fostering group unity in the face of

competition with other groups. We do not generally consider similarity of function to be at all a reason to assign a similar moral value, and for good reason. In the same way, the fact that cooperation with each other and a hypocritical self-representation both evolved for the same function has little relevance for how we will be constructing our ideal moral values. An equally troublesome consequence of the moral ambivalence of our history is that is that we cannot simply consult our nature to determine how we ought to live, because the average human tendency is not necessarily something we would want to encourage. Statistical tendency can be important in science but does not have to be morally important to us as individual humans. Even if psychology bears out an evolutionary prediction that the average person's motives are mixed, generosity is calculated, advice is self-serving, and modesty is false—even if this ends up being true, whoever said that goodness had to be the mean, had to be typical? That is the doctrine of a hopeful sort of humanist religion we have always had plenty of reason to doubt, long before Alexander and Trivers started talking about self-deception. The game theorists who consider nearly all of us “good guys”, and get their papers into *Science* when they promise that cheats really do lose out, owe much more to that fuzzy religion than to evolutionary biology. Likewise those ecologists for social justice, by warning against Alexander, were not defending society or justice so much as their own comforting conception of human nature. Seeing ourselves as honest and caring from the core of our beings, if this is a false view, will not help us make ourselves or the world better.

If the evolutionary account is correct about the moral ambivalence of our history and nature, we can sympathize with calls to widespread empathy (e.g., Baron-Cohen 2011), while realizing, with Darwin, that such an idea has absolutely no evolutionary precedent, and so must be taught (1871, ch. 4), and also realizing, with Alexander, that bringing about such empathy will likely be arduous because of contrary influences. If we do have widespread tendencies towards manipulation or false moral pretenses, most likely we will be in a much better position to rectify social ills if we are aware of these tendencies and can take them into account. This I take to be the import of the final sentiment of the associated paper, “a society of well-meaning people who understand themselves and their history very well is a better milieu than a society of well-meaning people who do not”. A bigger question is whether knowledge of the evolutionary basis of our psyche can enhance our ability to live up to whatever ideals or goals we do espouse: take “global cooperation” for instance. On this point, Richard Alexander is hopeful but realistic. In a paper he recently called his scientific swan song, he concluded by questioning whether the situation that started us competing in the first place—our genetic differences, and subsequent conflicts of interest—would ever subside enough in importance to permit peaceful coexistence as a species. As he exclaimed in the last sentence, “If only we could devise an effective way to tackle the question—and generate the solution—of how to make ‘otherness’ go away!” (Alexander 2012).

There are two complementary senses in which we can say that morality deals with the ideal: the first is, as Alexander said, that the moral is never actualized as it is conceived. The other spirit, the other side of the coin, is that the moral is something to which we can aspire if we so choose. By engaging in such encouraging words I fully admit to be falling right into Alexander's prediction that each individual will urge others to be more moral than oneself. Avoiding autopsychanalysis, if I accept for the sake of argument a degree of duplicity in my moralizing, what is the next step? Should I abandon the whole

project and be nihilistic? Or just keep my opinions to myself? An alternative strategy is to go with the flow in just this one area: what if I try to convince others to be better, and they try to convince me to be better, and I like a sucker fall for them rather than steel myself with cynicism. If we were to do this, society might remain a decent place to live, and might even become more so, whereas rejecting the evolutionary account or throwing up our hands in moral skepticism seem far less promising options. This is not blindness to the truth, it is merely rolling with the evolutionary punches in a case (like eating and breathing) where it is actually good for us to do so. We so often seem to be struggling to muffle or channel recalcitrant aspects of our evolutionary heritage; let's just relax in this particular area and allow ourselves to moralize and be taken in by moralizing. But of course you know there could be a devil on my shoulder when I say that.

### Acknowledgments

Many thanks to the editors for their helpful suggestions, and to Andrew Richards for extensive input and discussion.

### References

- Alexander, R. D. 1978: Evolution, creation, and biology teaching. In: *American Biology Teacher*. pp. 91-96, 101-104, 107.
- 1979: *Darwinism and Human Affairs*. University of Washington Press, Seattle.
  - 1982: Biology and the moral paradoxes. *Journal of Social and Biological Structures* **5**, 389-395.
  - 1987: *The Biology of Moral Systems*. Aldine de Gruyter, Hawthorne, NY.
  - 1988: Evolutionary approaches to human behavior: what does the future hold? In: *Human Reproductive Behavior: A Darwinian Perspective*. (Betzig, L. L., Borgerhoff Mulder, M. & Turke, P. W., eds). Cambridge University Press, Cambridge. pp. 317-341.
  - 1989: Evolution of the human psyche. In: *The Human Revolution*. (Mellars, P. & Stringer, C., eds). University of Edinburgh Press, Edinburgh. pp. 455-513.
  - 1990: How did humans evolve? Reflections on the uniquely unique species. *University of Michigan Museum of Zoology Special Publication* **1**, 1-38.
  - 1992: Biological considerations in the analysis of morality. In: *Evolutionary Ethics*. (Nitecki, M. H. & Nitecki, D. V., eds). SUNY Press, Albany, NY. pp. 163-196.
  - 2005: Evolutionary selection and the nature of humanity. In: *Darwinism and Philosophy*. (Hosle, V. & Illies, C., eds). University of Notre Dame Press, Notre Dame. pp. 424-495.
  - 2009: Understanding ourselves. In: *Leaders in Animal Behavior: The Second Generation*, vol. II. (Drickamer, L. C. & Dewsbury, D. A., eds). Cambridge University Press, New York.
  - 2012: Darwin's challenges and the future of human society. In: *Prediction: Breakthroughs in Science, Markets, and Politics*. (Wayman, F., Williamson, P. & Bueno de Mesquita, B., eds). University of Michigan Press, Ann Arbor, MI.
- Alexander, R. D. & Borgia, G. 1978: Group selection, altruism, and the levels of organization of life. *Annual Review of Ecology and Systematics* **9**, 449-474.
- Baron-Cohen, S. 2011: *The Science of Evil: On Empathy and the Origins of Cruelty*. Basic Books, New York.

- Batson, C. D., Thompson, E. R., Seuferling, G., Whitney, H. & Strongman, J. A. 1999: Moral hypocrisy: appearing moral to oneself without being so. *Journal of Personality and Social Psychology* **77**, 525-537.
- Brockmann, H. J. 2001: The evolution of alternative strategies and tactics. *Adv Stud Behav* **30**, 1-51.
- Burt, A. & Trivers, R. 2006: *Genes In Conflict: The Biology of Selfish Genetic Elements*. Harvard University Press, Cambridge.
- Churchland, P. S. 2011: *Braintrust: What Neuroscience Tells Us About Morality*. Princeton University Press, Princeton, NJ.
- Darwin, C. 1871: *The Descent of Man and Selection in Relation to Race*, 1874 edn. John Murray, London.
- de Waal, F. 1996: *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Harvard University Press, Cambridge, MA.
- Descartes, R. 1637: *A Discourse on Method*, 1969 edn. Everyman, Dent, London.
- Ghiselin, M. 1974: *The Economy of Nature and the Evolution of Sex*. University of California Press, Berkeley, CA.
- Gross, M. R. 1996: Alternative reproductive strategies and tactics: diversity within sexes. *Trends Ecol Evol* **11**, 92-98.
- Haidt, J. 2006: *The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom*. Basic Books, New York.
- Hobbes, T. 1651: *Leviathan*, 1839 edn. John Bohn, London.
- Hrdy, S. B. 2009: *Mothers and Others: The Evolutionary Origins of Mutual Understanding*. Belknap Press, Cambridge, MA.
- Keller, L. & Ross, K. G. 1998: Selfish genes: a green beard in the red fire ant. *Nature* **394**, 573-575.
- Kierkegaard, S. 1846: *Purity of Heart Is to Will One Thing*, 1939 edn. Harper, New York.
- Kurzban, R. 2010: *Why Everyone (Else) Is A Hypocrite: Evolution and the Modular Mind*. Princeton University Press, Princeton, NJ.
- Lahti, D. C. 2003: Parting with illusions in evolutionary ethics. *Biology and Philosophy* **18**, 639-651.
- Lahti, D. C. & Weinstein, B. S. 2005: The better angels of our nature: group stability and the evolution of moral tension. *Evol Hum Behav* **26**, 47-63.
- Murdoch, I. 1970: *The Sovereignty of Good: Studies in Ethics and the Philosophy of Religion*. Routledge and Kegan Paul, London.
- Oakley, B. 2007: *Evil Genes: Why Rome Fell, Hitler Rose, Enron Failed, and My Sister Stole My Mother's Boyfriend*. Prometheus Books, New York.
- Trivers, R. 2000: The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences* **907**, 114-131.
- Wittgenstein, L. 1921: *Tractatus Logico-Philosophicus*, 1974 edn. Routledge and Kegan Paul, London.