

4 The evolution of evaluators

Daniel C. Dennett

4.1 Whence cometh our values?

We have values and aspirations. What of other animals? Are their “values” different from ours? Animals manifestly prefer having plenty of food to starvation, comfort to pain, and they will work hard to obtain a mate. But beyond these “culture comforts,” they seem to be largely indifferent to the prospects and anxieties that make up human life. A suitable coverall term for human aspiration would be *the pursuit of happiness*, bearing in mind that happiness is many different things to different people. This already sets us aside from our fellow creatures. To put it vividly, Mother Nature doesn’t care whether we are happy – but we care (and Mother Nature doesn’t care that we care). That is, it would be naive to suppose that the process of natural selection has somehow endorsed our pursuit of happiness as a proximal mechanism for maximizing our genetic fitness. It is consistent with what we know of evolution to suppose that the process of natural selection – Mother Nature – would design us to experience however much anxiety and torment consistent with making more grandchildren. Our values are, like everything else in our extended phenotypes, products of evolutionary processes, but we misuse them if we see them to be just like the “values” of other animals, which can indeed be viewed as the straightforward result of Mother Nature’s project of installing an optimally reliable fitness-enhancing set of preferences. The difference arises, I would argue, from the fact that we have culture, and culture provides a medium in which a radically different – indeed, orthogonal – set of selection pressures can re-direct evolutionary processes into unprecedented channels.

There is an immense gap between us and other animals, even though we are all non-miraculous products of natural selection. Whence cometh our norms, our standards against which to judge decisions? Some of them, no doubt, are created by a sharpening and extension of our animal natures: like other creatures, we can distinguish comfort from discomfort, easy from difficult, safe from risky. If all our norms had such a genealogy, economics would not only be a lot easier; it would be a straightforward branch of ecology, human foraging theory, in effect. Economics is – or ought to be – more interesting than that, because human desire is more interesting than that. The “animal model” might be appropriate for considering

the welfare of severely retarded children in an institution, for instance, for they are apparently incapable of responding to, recognizing, or benefiting from the goods that matter so much to the rest of us. Keep them comfortable, well-fed and stimulated in whatever ways they care about, and that is about as good a life as they could have. It is a good life for a dog or a tiger or a chicken or a fish, but it is not a good life for a normal human being.

Consider an example drawn from a recent study of subjective well-being: "He believes that others believe that he is responsible for his unemployment." It is obvious that such a belief could play a major role in determining the (un-)happiness of a human being: it's baleful influence could overpower the positive contributions of comfort, food, health and plenty of sex. The *basis* for such an attitude might well be genetically laid down. It might, for instance, be grounded somehow in our species' history of living in groups ordered by dominance hierarchies, but in its convolution of nested attitudes it goes way beyond any factors that could play a role in the subjective well-being of any non-human creature.

Our genetic heritage gives us a biological base on which to build our values, but a base is only a base. It provides us with dispositions and preferences and organs that can then be adapted for other purposes, cobbling up ancient competences to new uses. We must not commit the genetic fallacy of assuming that we use these organs today for what they were evolved for. Consider, for instance the schematic graph in figure 1.

Let goodness be whatever you think goodness is, and look at the two life trajectories, A and B. The person on trajectory A gets off to a fine start, and has lots of goodness (whatever that is) for the rest of his life. The person on trajectory B has a life that steadily gets better, and eventually reaches the same level of goodness as on trajectory A, but the total goodness accumulated by the person on trajectory A is roughly twice as much. It would seem that life A should be preferred to life B, and yet many different studies suggest that no matter how you operationalize the positive value on the vertical axis, people tend to be happier when their lives are steadily getting better than when their lives are "stagnating" at a high level, no matter how high. Who says this is wrong? On what basis could it be demonstrated that a preference for trajectory B is irrational?¹ Is it "unnatural"? Maybe it is (on some construal of the term), but so what? Many of the things we prize are "unnatural." Consider yet another trajectory C, a roller-coaster ride of ups and downs. Any novelist will tell you that these are the lives that make the best stories, and who is to say that it would be irrational to prefer a life that makes a good story – even a tragedy – to a life of comfort, ease, and a plethora of descendants?

Who is to say? Not Mother Nature. Mother Nature doesn't care. Perhaps at bottom the reason people tend to prefer a life trajectory that gets better and better is that they have been designed by evolution to be better at detecting changes than at detecting absolute values, and they are designed to detect relevant (value-laden) changes more than others. The sorts of evaluations they are good at are "getting warmer" or "feeling better" or "hurting more" or "getting hungrier." If they come to value change for change's sake, this may well be, from Mother Nature's point of view, a mistake, but so what? Psychologists describe a personality type they

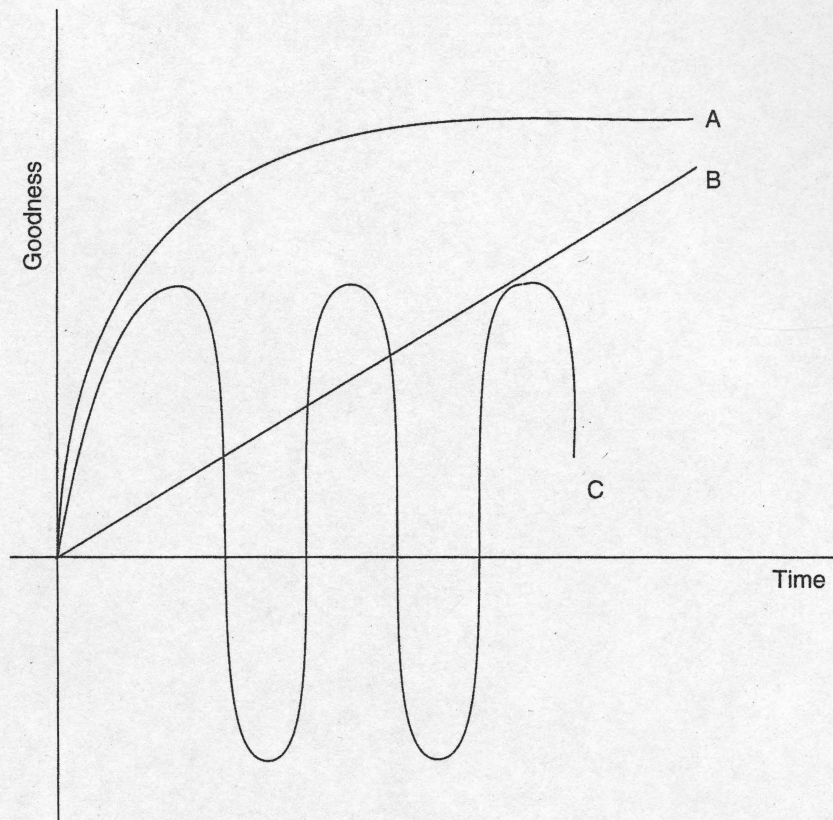


Figure 4.1

call “sensation seeker,” and they tend to impose their own values on the category by treating it as a mild pathology, a subnormal or aberrant type. But if some of us want to be sensation-seekers, why shouldn’t we have our heart’s desire? Why should we puritanically resolve to stick to the norms that guide the other species – or the clinical psychologists? In fact, we just don’t. Not only do we go our merry way, adopting a wide variety of values and life goals; we declare that very freedom to be a value.

As Cavalli-Sforza and Feldman (1981) note,

There are people determined to risk their life to reach the top of Mt. Everest, and others that spend their life accumulating money, or attempting artistic or scientific creations, or simply trying to do as little as possible. It is difficult to subsume all of these choices under a common schedule admitting no individual variation.

Control is delegated to a system of poorly understood internal drives and rewards that direct the activity of the individual, . . . (p.342)

they go on to say, and

Our very inadequate knowledge of this steering system prevents us from making finer statements, but it is probably true that the system's overall activity is directed towards maximizing self satisfaction of the individual. Important complications arise because we can satisfy ourselves in many different, competing ways, many of which demand careful advance planning (p.364).

These important complications are the result of structures we build in our decision-making systems, structures that incorporate our values, as well as our empirical knowledge. Since we are not born with them, they have to be installed, in the course of acculturation and learning. This process, which exploits the equipment Mother Nature gave us in ways she never "intended," might seem to lead to cultural relativism of the most nihilist kind. If culture is the source of the values that Mother Nature doesn't give us, isn't one culture as good as another? Does might make right, or does majority rule? Is there any objective standpoint from which some bit of social construction could ever be judged to be a social *misconstruction*?

The fear of a slippery slide into relativism should not panic us into endorsing the imperative of our genes. That would be to commit the genetic fallacy, just as surely as if we endorsed the norms of, say, the Bible. What was good enough for the reptiles or the apes, or good enough for the Samaritans or the Israelites, *may or may not* be good enough for us – that is an open question, which we must decide.

And we *can* decide, for good reasons or for bad reasons. How is it that we, alone among the creatures, have this choice point, this unprecedented opportunity to bootstrap ourselves beyond the norms of our biological heritage? Richard Dawkins closes his book, *The Selfish Gene*, with the ringing declaration:

We have the power to defy the selfish genes of our birth and, if necessary, the selfish memes of our indoctrination We are built as gene machines and cultured as meme machines, but we have the power to turn against our creators. We, alone on earth, can rebel against the tyranny of the selfish replicators (1976, p.215).

But how is this possible? Dawkins doesn't say. I think, however, that his brief, informal account of cultural evolution presents the key ideas needed for an explanation of this fact about human nature.

What Cavalli-Sforza and Feldman call the "steering system" is something that emerges over time in a maturing human being. It differs from culture to culture, and from individual to individual. We are all agents, but we are not all the same agent; we differ in our beliefs, and in our values – and in the way we think about values. It is important to ask if we can keep this last factor distinct from the others, since there is a natural tendency for differences in ways of thinking to disappear in our models. When two agents agree on the *summum bonum*, and on the ranking of their subsidiary goals, and on the background information that is relevant, can they arrive at different courses of action? Suppose they have all the same beliefs and desires (suppose, in my terminology (Dennett 1971, 1987), that they are identical

intentional systems). To a first approximation, at least, we predict that they will make the same decisions, choose the same actions, because they will choose to do whatever it is they deem to be most rational, given their beliefs and desires – which *ex hypothesi* are identical. This assumes that there is just one canon of rationality, optimal decision procedure. And this assumption can be maintained, come what may, by recasting any apparent differences in canons of rationality that emerge as differences in beliefs or values. I am not deploring this possibility; just reminding ourselves of it. It may be wise to idealize to a content-free framework of rationality and treat all differences as differences in beliefs and desires, information and goals.

Some such differences are more fundamental than others, of course. Consider, for instance, a question of paramount importance: *Cui bono?* Who benefits? (Dennett 1995a: 324–330) When an agent or intentional system makes a decision about which is the best course of action, all things considered, we need to know from whose perspective this optimality is being judged. A more or less standard default assumption, at least in the Western world, and especially among economists, is to treat the agent as a sort of punctate, Cartesian locus of well-being. What's in it for *me*? Rational *self-interest*. But while there has to be something in the role of the self – something that defines the answer to the *Cui bono* question for the decision-maker under examination, there is no necessity in this default treatment, common as it is. A self-as-ultimate-beneficiary can in principle be indefinitely distributed. I can care for others, or for a larger social structure, for instance. There is nothing that restricts me to a *me* as contrasted to an *us*. (This is distinct from Bowles' and Gintis' concept of community [this volume]; I am saying that, independently of any particular history of individuals or structure of community it is an open possibility that a mind – the “steering system” that gets built in a human brain – treat something other than the well-being of the body it inhabits as the touchstone of value.)

One tradition would speak here of “selfless” caring, but since this inevitably invites cavils about the purported incoherence of true selflessness, I prefer to think of this as the possibility of extending the domain of the self. I can still take my task to be looking out for #1 while including, under #1, not just myself, but my family, the Chicago Bulls, Oxfam Here is one good reason for treating the self this way: Suppose I am an agent in a bargaining situation, or in a prisoner's dilemma, or faced with a coercive offer, or an attempt at extortion. My problem is not resolved, or diminished, or even significantly adjusted, if the “self” I am protecting is other than my proper self, if I am not just trying to save my own skin so to speak. An extortionist or a benefactor who knows what I care about is in position to frame the situation to hit me where it matters to me, whatever matters to me.

We human beings can distribute our selves not only in “space” but also in time, caring about our own futures and pasts, and even about remote pasts and futures centuries outside our individual lifetimes. No other species has such labile boundaries on the problems of self-interest its members can define. And it is worth noting that we are quite properly concerned with the prospect of this lability moving in the other direction. As I put it in *Elbow Room*, in my discussion of the Incredible Disappearing Self, “If you make yourself really small, you can externalize almost

everything.” (1984: 143) Shrinking the boundaries of the self can be a way of obtunding suffering (as when an abused child dissociates and “leaves” [Dennett, 1995b]), but it can also be a way of evading responsibility. The standard Western model of the self is not as universally appreciated as we in the West often suppose. “To us in Asia,” says Lee Khan Yew, Senior Minister of Singapore, “an individual is an ant. To you, he’s a child of God; it is an amazing concept.” (*Boston Globe*, April 29, 1994) Recognizing that this statement is itself a piece of propaganda, not remotely an accurate scientific observation, only drives home the point that selves are socially constructed (and none the less real and objective on that account), and hence pre-eminently cultural entities, not simply part of our biological equipment.

4.2 Perspectives on cultural evolution

How does culture accomplish the design work in our brains that makes all these transformations possible? Partly by restructuring the functional structure of individual human brains – creating virtual machines on the underlying hardware (for more on the mind as a virtual machine, see Dennett 1991, 1996), but also by creating structures in the public world that alter the perceivable opportunities and costs for those virtual machines. For instance, consider Bowles’ definition of a norm [this volume]. “A norm is a cultural trait governing actions that affect the well-being of others but that cannot be regulated by costlessly enforceable contracts.” By making norms, culture makes habits; this makes it easier for brains to be the kinds of minds they are. As Andy Clark has put it, “We use intelligence to structure our environment so that we can succeed with less intelligence. Our brains make the world smart so we can be dumb in peace!” (1996: 180)

If culture transmits and installs virtual machines in growing human brains and their surroundings, where does the software come from? Who writes the code? Nobody. Who invented Chinese? Who invented arithmetic? Who invented money? Nobody. These fine artifacts, all exhibiting impressive features that bespeak expensive histories of R and D, have evolved over long periods of time, the design work distributed among myriads of largely oblivious innovators and editors.

When one says that cultures evolve, this can be taken as a truism, or as asserting one or another controversial, speculative, unconfirmed theory. Consider a cultural inventory at time *t*: it includes all the languages, practices, ceremonies, edifices, methods, tools, myths, music, art, and so forth, that compose a culture. Over time, the inventory changes. Some items disappear, some multiply, some merge, some change. (When I say some change, I mean to be neutral at this point about whether this amounts to their being replaced by similar items, or their undergoing a transformation.) A verbatim record of this history would not be science; it would be a data base. That is the truism: cultures evolve over time. Now the question remains: how are we to explain the patterns found in that data base? Are there any good theories or models of cultural evolution?

The traditional model to be found in most accounts by historians and anthropologists treats culture as composed of goods, possessions of the people, who husband them in various ways, wisely or foolishly. They carefully preserve their

traditions of fire-lighting, house-building, speaking, counting, justice, etc. They trade cultural items as they trade other goods. And of course some cultural items (wagons, pasta, recipes for chocolate cake, etc.) are definitely goods, and we can plot their trajectories using the tools of economics. The people, on this model, are seen as having an autonomous or independent rationality; deprive a person of goods, and he stands there, naked but rational and full of informed desires. When he clothes himself and arms himself and equips himself with goods, he increases his powers, complicates his desires, etc.

On this way of thinking, the relative "replicative" power of various cultural goods is measured in the marketplace of cost-benefit calculations performed by the people. If Coca-Cola bottles proliferate around the world, it is because more and more people prefer to buy a Coke. Advertising may fool them. But then we look to the advertisers, or those who have hired them, to find the relevant loci of values for our calculations. *Cui bono?* The purveyors of the goods, and those they hire to help them.

Biologists, too, can often make sense of the evolution (in the neutral sense) of features by treating them as goods: one's food, one's nest, one's burrow, one's territory, one's mate[s], one's time and energy. Cost-benefit analyses shed light on the husbandry engaged in by the members of the different species inhabiting some shared environment. Not every "possession" is considered a good, however: one's accompanying flies and fleas, the dirt and grime that accumulates on one's body, are of no value, or of negative value, for instance. One's symbionts are not normally considered as goods by biologists, except when the benefits derived from them (by *whom?*) are manifest.

This perspective is not uniformly illuminating, nor is it obligatory. I would like to suggest that both biologists and economists (and other social scientists) can benefit from adopting a different vantage point on these phenomena, one which quite properly gives pride of place to the *Cui bono* question, which can provide alternative answers that are often overlooked. This is Dawkins' meme's-eye point of view, which recognizes – and takes seriously – the possibility that cultural entities may evolve according to selectional regimes that make sense only when the answer to the *Cui bono* question is that it is the cultural items *themselves* that benefit from the adaptations they exhibit.

Dawkins' theory of memes, as briefly sketched in a single chapter of *The Selfish Gene* (1976, but see also Dawkins 1993), is hardly a theory at all, especially compared to the models of cultural evolution developed by other biologists, such as Cavalli-Sforza and Feldman (1981), Lumsden and Wilson (1981), and Boyd and Richerson (1985). Unlike these others, Dawkins offers no formal development, no mathematical models, no quantitative predictions, no systematic survey of relevant empirical findings. But Dawkins does present an idea that is overlooked by all the others, and it is, I think, a most important idea. It is the key to understanding how we can be not just guardians and transmitters of culture, but cultural entities ourselves – all the way in.

Whenever costs and benefits are the issue – e.g., when Pagano [this volume speaks of how a difference in the distribution of property rights leads to a difference

in accounting the profits, or when Bowles [this volume] speaks of “group beneficial effects” and of the “efficiency-enhancing properties” of a practice, we need to ask *Cui bono*? A benefit by itself is not explanatory; a benefit in a vacuum is indeed a sort of mystery; until it can be shown how the benefit actually redounds to enhance the replicative power of a replicator, it just sits there, alluring, perhaps, but incapable of explaining anything.

We see an ant laboriously climbing up a stalk of grass. Why is it doing that? Why is that adaptive? What good accrues to the ant by doing that? That is the *wrong question* to ask. No good accrues to the ant; its brain has been invaded by a fluke (*Dicrocoelium dendriticum*), one of a gang of tiny parasites that need to get themselves into the intestines of a sheep in order to reproduce (Ridley 1995: 258). (Salmon swim up stream, these parasitic worms drive ants up grass stalks, to improve their chances of being ingested by a passing sheep.) The benefit is not to the reproductive prospects of the ant but the reproductive prospects of the fluke.

Dawkins points out that we can think of cultural items, memes, as parasites, too. Actually, they are more like a simple virus than a worm. Memes are supposed to be analogous to genes, the replicating entities of the cultural media, but they also have vehicles, or phenotypes; they are like not-so-naked genes. They are like viruses (Dawkins 1993). As with viruses, there is a phenotype/genotype distinction, but just barely. Basically, a virus is just a string of DNA (or RNA) with attitude. And similarly, a meme is an information-packet (the information, not the vehicle) with attitude – with some phenotypic clothing that has differential effects in the world that thereby influence its chances of getting replicated.

And in the domain of memes, the ultimate beneficiary, the beneficiary in terms of which the final cost-benefit calculations must apply is: the meme itself, not its carriers. This is not to be read as itself a bold empirical claim, ruling out (for instance) the role of individual human agents in devising, appreciating and securing the spread and prolongation of cultural items. It is rather a proposal that we adopt a perspective or point of view, from which a *wide variety of different* empirical claims can be compared, and the evidence for them considered in a neutral setting, a setting that does not prejudge these hot-button questions.

In the analogy with the fluke, we are invited to consider a meme as like a parasite which commandeers an organism for its own replicative benefit, but we should remember that symbionts can be classified into three fundamental categories:

- *parasites*, whose presence lowers the fitness of their host;
- *commensals*, whose presence is neutral (though, as the etymology reminds us, they “share the same table”); and
- *mutualists*, whose presence enhances the fitness of both host and guest.

Since these varieties are arrayed along a continuum, the boundaries between them need not be too finely drawn; just where benefit drops to zero or turns to harm is not something to be directly measured by any practical test, though we can explore the consequences of these turning points in models.

The main point to note is that we should expect memes to come in all their varieties, too. This means, for instance, that it is a mistake to *assume* that the “cultural selection” of a cultural trait is always “for cause” – always because some perceived (or even misperceived) benefit it provides to the host. We can always *ask* if the hosts, the human agents that are the vectors, perceive some benefit and (for that reason, good or bad) assist in the preservation and replication of the cultural item in question, but we must be prepared to entertain the answer that they do not. In other words, we must consider as a real possibility the hypothesis that the human hosts are, individually or as a group, either oblivious to, or agnostic about, or even positively dead set against, some cultural item, which nevertheless is able to exploit its hosts as vectors.

The most familiar cases of cultural transmission and evolution discussed are innovations that are obviously of some direct or indirect benefit to the *Darwinian* – that is, genetic – fitness of the host. A better fishhook catches more fish, feeds more bellies, makes for more surviving grandchildren, etc. The only difference between stronger arms and a better fishhook in the (imagined) calculation of impact on fitness is that the stronger arms might be – might be – passed on quite directly through the germ line, while the fishhook definitely must be culturally transmitted. (The stronger arms could be culturally transmitted as well, of course. A tradition of body-building, for instance, could explain why there was very *low* [genetic] heritability for strong adult arms, and yet a very *high* rate of strong adult arms in a population.) But however strong arms or fishhooks are transmitted, they are typically supposed to be a good bargain from the perspective of genetic fitness. The bargain might, however, be myopic – only good in the short run. After all, even agriculture, in the long run, may be a dubious bargain if what you are taking as your *summum bonum* is Darwinian fitness. What alternatives are there?

First, we need to note that in the short run (evolutionarily speaking – that is, from the perspective of a few centuries or even millennia) something might flourish independently of whether it was of *actual* benefit to genetic fitness, but is strongly linked to whether it was of *apparent* benefit to genetic fitness. Even if you think that Darwinian fitness enhancement is the principle driving engine of cultural evolution, you have to posit some swifter, more immediate mechanism of retention and transmission.² It’s not hard to find one. As I noted earlier, cultural items may exploit machinery that had earned its keep in the past by embodying “fitness-enhancing set of preferences.” We are genetically endowed with a qualitative space in which some things feel good and some things don’t, and we tend to live by the rule: if it feels good, keep it. This rough and ready rule can be tricked, of course. The sweet tooth is the standard example. The explosion of cultural items – artifacts, practices, recipes, patterns of agriculture, trade routes – that depend quite directly on the exploitation of the sweet tooth has probably had a considerable net negative effect on human genetic fitness. Notice that explaining the emergence of these cultural items by citing their “apparent” benefit to genetic fitness does not in any way commit us to the (preposterous) claim that people *think* (mistakenly) that they are enhancing their genetic fitness by acquiring and consuming sugar. The rationale is not their’s, but Mother Nature’s. They just go with what they like.

Still, given what they like, they choose rationally, and indeed ingeniously and often with impressive foresight, how to obtain what they like. This is still the traditional model of cultural evolution, with agents husbanding their goods in order to maximize what they prefer - and getting their preferences quite directly from their genetic heritage. A more interesting possibility is acquiring new preferences that are themselves culturally transmitted symbionts of one sort or another. Each will have to bootstrap itself into the memosphere by exploiting some pre-established preference, but this recursive process, which can proceed at breakneck speed relative to the glacial pace of genetic evolution, can transform human agents indefinitely far away from their genetic beginnings. In an oft-quoted passage, E.O. Wilson claimed otherwise:

The genes hold culture on a leash. The leash is very long, but inevitably values will be constrained in accordance with their effects on the human gene pool.

(Wilson 1978: 167)

This leash. I am claiming, is indefinitely long, in the sense that the constraints Wilson speaks of can be so co-opted, exploited, and obtunded in a recursive cascade of cultural products and meta-products that it is not clear that there are *any* points in imaginable cultural design space that could not, in principle, be occupied by some product that could ultimately be traced back, via Wilson's leash of historical processes, to the genes. Many of these imaginable points would no doubt be *genetic cul-de-sacs* (*H. sapiens* would sooner or later go extinct as a result of occupying those points), but this is no barrier to their evolving in the swift time of cultural history.

Not only can we acquire tastes; we can acquire meta-tastes. That is, we can discover in the culture, and thereupon adopt, a taste for "cultivating" further acquired tastes, and so forth. At each stage we can anticipate finding parasites, commensals and mutualists - but we can classify these only by asking the *Cui bono?* question *against a new background* and making one local determination or another. One person's scholarly connoisseurship is another person's addiction to trash. Meta-memes for "traveling" or "being a collector" or "having a hobby" or "educating oneself" can themselves be viewed as either exploiters or enhancers of the pre-established *personal* (no longer genetic) preferences. It is interesting that in common parlance we often call our preferences "weaknesses," - as in "I have a weakness for strong cheese (or puns or redheads)" - deftly implying a standard to which in the same breath we deny any personal allegiance.

And this, then is the main point I wanted to emphasize in Dawkins' vision. The memes that proliferate will be the memes that replicate *by hook or by crook*. Think of them as entering the brains of culture members, making phenotypic alterations thereupon, and then submitting themselves to the great selection tournament - *not* the Darwinian genetic fitness tournament (life is too short for that) but the Dawkinsian meme-fitness tournament. It is *their* fitness as memes that is on the line, not their host's genetic fitness, and the environments that embody the selective pressures that determine their fitness are composed in large measure of *other* memes.

Why do their hosts put up with this? Why should the overhead costs of establishing a whole new system of differential reproduction be borne by members of

H. sapiens? Note that the question to be asked and answered here is parallel to the question we ask about any parasite-host relationship: why do the hosts put up with it? And the short answer is that it is too costly to eradicate, but this just means that the benefits accruing to the machinery that is being exploited by the parasites are so great that keeping the machinery and tolerating the parasites (to the extent that they are tolerated) has so far been the best deal available. And whether or not in the long run (millions of years) this infestation will be viewed as mutualism, commensalism or parasitism, in the short run (the last few millennia) the results have been spectacular: the creation of a new biological type of entity: a person.

I like to compare this development to the arrival of the eukaryotes seven billion years ago. Relatively simple prokaryotes got invaded by some of their neighbors, and the resulting endosymbiotic teams were more fit, and prospered, enabling a biological revolution. The eukaryotes, living alongside their prokaryotic cousins, but enormously more complex, versatile and competent, opened up a new design space of multi-cellular organisms. Similarly, the emergence of culture-infected hominids has opened up yet another region of hitherto unoccupied and untraversable design space. We live alongside our animal cousins, but we are enormously more complex, versatile and competent. And by joining forces with our memes, we create new candidates for the locus of benefit, new answers to the question: *bono?*

4.3 Some paths not taken

In Cavalli-Sforza and Feldman's pioneering work on cultural evolution, they note the phenomena that invite the meme's-eye view, but treat them as complicated and best set aside. As noted before, they discuss what they call the "steering system" and observe that "it is probably true that the system's overall activity is directed towards maximizing self satisfaction of the individual" (1981: 364). But they do go on to look at the possibilities this opens up.³ They briefly consider the prospect of treating artifacts, such as violins and cars, as "second order organisms," and measuring *their* fitness as "cultural objects" (p.17), but they do not suggest that such fitness might be anything other than excellence of design *from the point of view of artifact-user*, which is, as we have seen, just one of the possibilities – the analogue of the mutualist case, in effect. Thus when they speak of adaptation, they apparently have in mind only the genetic fitness of members of human cultures. For instance: "If cultural innovations are not truly random, but are designed to solve specific problems, they may increase the rate of the corresponding adaptation in evolution over that expected for a truly random process" (p.66). They go on to see that the chance that an intended improvement will be "truly adaptive in the long run is not 100%" so "a significant proportion of new cultural mutations might be truly random without any semblance of adaptiveness." But many of the mutations in this "significant proportion" might exhibit clear adaptiveness, measured from the meme's-eye perspective.

This overlooked opportunity is compounded by another. Cavalli-Sforza and Feldman correctly draw our attention to the distinction between what they call

awareness and adoption. Awareness is the minimal result of exposure; adoption involves a change of phenotype as a result of that exposure. The distinction is analogous to that between testing positive for a virus and having the full-blown symptomatology of the viral disease. The growth of awareness of a meme can be much swifter than its adoption, of course. The adoption/awareness distinction is important, but by restricting their models to the spread of adoption, they submerge a major channel of cultural evolution: "However, the final test of fitness is whether the learned trait will be really incorporated into the final permanent phenotype of the individual, or alternatively forgotten, rejected or replaced. In the latter cases the trait evidently does not pass the test of cultural selection" (p.66).

Certainly memes can be widely transmitted without being "adopted." In Dawkins' terms one might say that a meme can have a limited phenotype and an "extended" phenotype. Beavers in captivity may not get to build dams, the most distinctive feature of their extended phenotype, but still they may reproduce and hence evolve. Similarly, a meme "in captivity" may reproduce without having its extended phenotype effect. For instance, Marco Polo brought the pasta meme from China to Italy in his mind-zoo; he didn't have to adopt the meme; he didn't have to become a pasta-maker. I have the idea of cannibalism, but am not a cannibal. I keep the meme alive, however, and can pass it on to others by talking about cannibalism. Many effects of cultural evolution depend on such silent transmission, transmission with negligible phenotypic effects in the carriers – aside from the all-important effects of *informing* the carriers on that topic and fostering further transmission. (A fitness-enhancing phenotypic trait for a meme is *being an interesting topic of conversation*, but it is shared by very many widely diverse memes.) By defining the presence of the cultural item in terms of the manifestation of the trait in individual phenotypes, Cavalli-Sforza and Feldman obscure these routes of cultural transmission, and hence overlook many large-scale phenomena of cultural evolution.

With the advent of telecommunications it seems reasonable to suggest that, for many socio-behavioral traits, the teacher-leader type of transmission has come to play a dominant role. Today the audience of a single social-cultural leader can be more than continental in size. For traits under this sort of influence, evolution, as we have seen, becomes extremely fast. Since the awareness increases so quickly via the mass media, it is the rate of acceptance that limits the rate of evolution of these traits (p.354).

This is one sort of case, but just as salient, one would think, are the cases that don't involve "teacher-leaders" as their source. The Heaven's Gate suicidal cult memes recently took full advantage of the telecommunications media, and will now lie, more or less dormant in millions, perhaps billions, of minds, where they can evolve in all manners of ways with very little adoption.

It may be true that these phenomena do not really represent overlooked opportunities for science after all. Perhaps they are not paths that have been ignored, but rather stony ground on which no science could be grown.⁴ Perhaps, as I have suggested (1995a, pp. 35–60), memetics cannot be turned into serious science

for the relatively boring reason that the requisite data-gathering cannot be (If fossils were not formed and preserved in sediment, and if DNA could not be sequenced, evolutionary biology would be very largely speculative, for similar tedious reasons.) But then at least we should acknowledge that the phenomena of cultural evolution are not exhausted by the few cases for which mathematical models of some realism can be constructed. And we shouldn't take those as evidence for the ubiquity of the constraints that make up Wilson's imagination. Suppose it were a fact that the only aspects of cultural evolution that can be mathematically modeled "are constrained in accordance with their effects on the human gene pool"; it wouldn't follow that all important phenomena of cultural evolution are similarly constrained.⁵

A confusion that misdirects the imagination of theorists in another direction derives, I suspect, from a subtle misreading of Darwin's original use of artificial selection (deliberate animal breeding) and "unconscious" selection (the unwitting promotion of favored offspring of domesticated animals) as bridges to his concept of natural selection.⁶ While it is true that Darwin wished to contrast the lack of foresight or intention in natural selection with the deliberate goal-seeking of the artificial selectors, in order to show how the natural process could in principle proceed without any mentality at all, he did not thereby establish (as it may seem to have supposed) that deliberate, goal-directed, intentional selection is not a subvariety of natural selection! The short legs of dachshunds, and the huge udders of Holsteins are just as much products of natural selection as the wings of the eagle; they just evolved in an environment that included a particularly well-focused selective pressure consisting of human agents. These phenotypes fall under the same laws of transmission genetics, the same replicator dynamics, as any other – as special and extreme cases in which the default "randomness" or noisiness of selective pressure has been greatly reduced.

Applied to cultural evolution, the implication is this: There is no conflict between the claim that artifacts (including abstract artifacts – memes) are the products of natural selection, and the claim that they are (often) the foreseen, designed products of intentional human activity. It appears that some thinkers in the new emerging school of evolutionary archeology have made this mistake. According to a critique by Boone and Smith (forthcoming), at least some evolutionary archeologists think that the only way to be hardheaded and scientific about the Darwinian evolution of culture is to deny all intention, all rationality, on the part of human culture-makers. They opt for "selection rather than decision-making" (p. 11). That is simply a mistake, for the same reason it would be a mistake to say that the fan plumage of prize pigeons is the result of decision-making *rather than* selection. Boone and Smith fall in the same trap, in their discussion of the interesting phenomenon of the spread of snowmobiles among the Cree in northern Canada. They are surely right that the adoption of snowmobiles by the Cree cannot be accounted for in terms of the differential biological replication of the snowmobile users, but they misread the more interesting meme's-eye view perspective. They say:

The alternative that 'snowmobile memes' were transmitted more effectively

than 'snowshoe memes' to non-descendant Cree (as well as offspring), while plausible, *is not natural selection* [emphasis added]; more significantly, it requires precisely the kind of adaptive decision-making that EA [evolutionary archeology] is dedicated to eliminating from archeological explanation [ms p.12].

On the contrary, if you adopt the meme's-eye perspective, in which the snowmobile meme is seen as the replicator, with *its own* fitness, then cultural evolution can be seen to be due to "adaptive decision-making" while also a variety of natural selection. Consider the fitness of the domesticated horses that spread so quickly among the Native Americans after their introduction, but then more recently, after the advent of the automobile, have dwindled sharply. These fluctuations in genetic fitness have been due to changes in the selective forces arrayed in the various environments in which the horses have existed, of course, and the fact that conscious, foresightful human agents form the key component in those selective environments does nothing to remove the phenomena from the domain of standard genetic evolution by natural selection.

Among those who have overlooked this fact is Steven Pinker, who dismisses models of cultural evolution in a brief passage in *How the Mind Works* (1997):

Stop being so literal-minded! respond the fans of cultural evolution. Of course cultural evolution is not an exact replica of the Darwinian version. In cultural evolution, the mutations are directed and the acquired characteristics are inherited. Lamarck, while being wrong about biological evolution, turned out to be right about cultural evolution. . . . To say that cultural evolution is Lamarckian is to confess that one has no idea how it works. The striking features of cultural products, namely their ingenuity, beauty, and truth (analogous to organisms' complex adaptive design), come from the mental computations that "direct" – that is, invent – the "mutations," and that "acquire" – that is, understand – the "characteristics" (p.209).

Pinker has imputed the wrong parallel; it is not Lamarck's model, but Darwin's model of *artificial* selection (as a special case of natural selection) that accommodates the phenomena he draws t out attention in this passage. And it is ironic that Pinker overlooks this, since the cultural phenomena he himself has highlighted as examples of evolution-designed systems, linguistic phenomena, are almost certainly *not* the products of foresightful, ingenious, deliberate human invention. Some designed features of human languages are no doubt genetically transmitted, but many others – such as changes in pronunciation, for instance – are surely culturally transmitted, and hence products of cultural, not genetic, evolution.

Some memes are like domesticated animals; they are prized for their benefits, and their replication is closely fostered and relatively well understood by their human owners. Some memes are more like rats; they thrive in the human environment in spite of being positively selected against – ineffectually – by their unwilling hosts. And some are more like bacteria or other viruses, commandeering aspects of human behavior (provoking sneezing, for instance) in their "efforts" to propagate from host to host. There is artificial selection of "good" memes – like the

memes of arithmetic and writing, which are carefully taught to each new generation. And there is unconscious selection of memes of all sorts – like the sub-mutations in pronunciation that spread through linguistic groups, presumably with some efficiency advantage, but perhaps just hitchhiking on some quirk of human preference. And there is unconscious selection of memes that are positively maladaptive, but which prey on flaws in the human decision-making apparatus, as provided for in the genome and enhanced and adjusted by other cultural innovations – such as the abducted-by-aliens meme, which makes perfect sense when *its own* fitness as a cultural replicator is considered.

4.4 Conclusion

This spectrum of possibilities, from unwitting, unconscious hosting of culture-borne viruses (of all “attitudes”) to the foresightful design and promulgation of inventions and creations that intelligently and artfully draw upon well-understood cultural resources, *must* be viewable under a single, unifying perspective. It is only from such a perspective that we can make sense of the trajectories that have taken us – and only us – beyond the horizons of our selfish genes, by creating new environments of selection – persons and their projects – that in evolution does not *deny* the possibility of moving to what might be called a *mind’s-eye* perspective of evaluation; it is precisely what makes such a transition – without any help from skyhooks – possible.

Notes

- 1 Perhaps we just haven’t yet found the “right” definition of goodness. But if we are constraining our attempt at definition by the demand that it yield a preference for trajectory A over trajectory B in “normal people,” then we are prejudging the issue in a way that itself requires motivation. Who says this is a constraint we should honor?
- 2 This is parallel to the familiar myopia of genetic evolution; the fact that sex, say, is good for maintaining genetic versatility in the long run is no explanation of how and why it gets maintained in the short run. The heavy and immediate costs of meiosis have to be balanced by some heavy and immediate benefit (see e.g., Ridley 1993).
- 3 “Coca-Cola, frisbee, volleyball and yo-yos are examples of ‘innovations’ that have spread rapidly through whole countries or continents. It is obvious that in none of these examples does participation appreciably alter the probability of surviving or having children” Cavalli-Sforza and Feldman (p.15).
Boyd and Richerson, similarly, make it clear that “Nonadaptive, or even frankly maladaptive, cultural variants can spread in a population under the influence of indirect bias, even in the face of selection and direct bias favoring more adaptive variants” (p.279). They do not, however, attempt to explore these possibilities.
- 4 In later work, Feldman and his colleagues have addressed some of the complexities they deliberately submerged in their earlier work. In a paper submitted to the Siena workshop, Feldman, Otto and Christiansen (1996) discuss models that have the potential to distinguish awareness (“a latent factor that can be transmitted culturally from generation to generation”) and adoption (“phenotypic effects”) [ms, p.12].
- 5 Aaron Lynch (1991) has attempted a mathematical formulation of some aspects of memetics, focusing on the simplest stripped down cases – which still swiftly spawn complications. Like Cavalli-Sforza and Feldman, he postpones treatment of the awareness

side of the awareness/adoption distinction, though his model permits the distinction to be expressed. Whether these tactical simplifications permit the resulting model to tell us anything surprising (and confirmed) about the real phenomena remains to be seen.

6 The following paragraphs are drawn, with revisions, from Dennett, forthcoming.

References

- Boone, James L. and Eric Alden Smith (forthcoming) A Critique of Evolutionary Archeology, *Current Anthropology*. [special issue, supplement, June 1998].
- Bowles, Samuel and Herbert Gintis (2000). Chapters 14 and 15, this volume
- Boyd, Robert and Peter J. Richerson (1985) *Culture and the Evolutionary Process*. Chicago and London: Chicago University Press.
- Cavalli-Sforza, Luca and Marcus Feldman (1981) *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton, NJ: Princeton University Press.
- Clark, Andy (1996) *Being There*, Cambridge, MA: MIT Press.
- Dawkins, Richard (1976) *The Selfish Gene*. (2nd edn, 1989). Oxford: Oxford University Press.
- (1993) Viruses of the mind. In Bo Dahlbom (ed.) *Dennett and his Critics*. Oxford: Blackwell.
- Dennett, Daniel (1971). Intentional systems. *Journal of Philosophy*, LXVIII: 87–106.
- (1984) *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press, and Oxford: Oxford University Press.
- (1987) *The Intentional Stance*. Cambridge, MA: Bradford Books/MIT Press.
- (1991) *Consciousness Explained*. Boston and New York: Little Brown.
- (1995a) *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster.
- (1995b) Animal consciousness: what matters and why. *Social Research* 62(3): Fall 1995, 691–710.
- (1996) *Kinds of Minds*. New York: Basic.
- (forthcoming) Snowmobiles, horses, rats, and memes: a comment on *A Critique of Evolutionary Archeology* by James L. Boone and Eric Alden Smith. *Current Anthropology*, [special issue, supplement, June 1998].
- Feldman, Marcus, Sarah P. Otto and Freddy B. Christiansen, Genes, Culture and Inequality. Submitted to the Siena workshop.
- Lumsden, C. and E.O. Wilson (1981) *Genes, Mind and Culture*, Cambridge, MA: Harvard University Press.
- Lynch, Aaron (1991) Units, events and dynamics of meme replication. *Journal of Ideas* [available in revised form at <http://www/mcs/net/aaron/mememath.htm>]
- Pagano, Ugo (2000) Chapter 2, this volume
- Pinker, Steven (1997) *How the Mind Works*. New York: Norton.
- Ridley, Mark (1995) *Animal Behaviour* (2nd edn), Boston: Blackwell Science.
- Ridley, Matt (1993) *The Red Queen: Sex and the Evolution of Human Nature*. London: Penguin.
- Wilson, E.O. (1978) *On Human Nature*. Cambridge, MA: Harvard University Press.