

DRAFT: Memes paper for Fyssen

July 12, 2000

From Typo to Thinko:

When Evolution Graduated to Semantic Norms

"Natural selection edits with an eye only toward what the message says, not to what it means." Richard Powers, *The Gold Bug Variations*, p546

1. Darwinian Perspectives on Culture

Do we need a Darwinian theory of cultural evolution? In one sense, certainly. It is obvious that there are patterns of cultural *change*-evolution in the neutral sense-and any theory of cultural change worth more than a moment's consideration will have to be Darwinian in the minimal sense of being *consistent with* the theory of evolution by natural selection of *Homo sapiens*. Our species name is well chosen, and it is culture that makes us the knowing hominid, so a minimally Darwinian theory of culture must hold that the phenotypic traits that make cumulative culture possible-mainly, language and the habits of sociality-evolved by natural selection, unaided by what I call *skyhooks*: saltations in Design Space that could not be the outcome of standard evolutionary mechanisms (Dennett, 1995). This minimal Darwinism is simply the denial of the hypothesis that culture is, as it were, a miracle, a gift from God.⁽¹⁾ It maintains in one way or another that natural selection eventually provided the foundations for culture, which then took off, and elaborated itself under some regime that explains the patterns of cultural change, but that regime need not itself be Darwinian in any interesting sense.

For instance, the standard model is an economic one: the theorist says, in effect, if Darwin will give me *Homo economicus*, a social group of rational, self-interested individuals getting and spending, saving and making and trading, I can then use the intentional stance (Dennett, 1971, 1987) as the explanatory framework for describing and accounting for the patterns of cultural evolution. This economic model is used not just by economists, of course; it is tacitly presupposed by historians and anthropologists and all the other theorists who treat culture as composed of goods, possessions of the people, who husband them in various ways, wisely or foolishly. People 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 so we can plot their trajectories using the tools of economics. It is clear from this perspective that highly prized cultural entities will be protected at the expense of less favored cultural entities, and there will be a competitive market where agents both "buy" and "sell" cultural wares. If a new method of house-

building or farming or a new style of music sweeps through the culture, it will be because people perceive advantages to these novelties and choose them. 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 agents whose desires fix the values for our cost-benefit calculations. *Cui bono?* Who benefits? The purveyors of the goods, and those they hire to help them. etc. On this way of thinking, then, the relative "replicative" power of various cultural goods--whether Coke bottles, building styles or religious creeds--is measured in the marketplace of cost-benefit calculations performed, consciously or unconsciously, by the people.

Biologists, too, make good use of the economic model, explaining the evolution (in the neutral sense) of features of the natural world by treating them as goods belonging to various members of various species: one's food, one's nest, one's burrow, one's territory, one's mates, 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.⁽²⁾ Not every "possession" is considered a good, however. The dirt and grime that accumulates on one's body, to say nothing of the accompanying flies and fleas, are of no value, or of negative value, for instance. These hitchhikers are not normally considered as goods by biologists, except when the benefits derived from them are manifest.

These economic models of culture are consistent with Darwinism but not more specifically Darwinian. Darwinian evolution provides organisms whose ultimate goal is self-replication, and who then track the rational if myopic trajectory to that end; the interactions of such rational agents determines which features of the shared environment will proliferate, which be contested, which despoiled, and so forth. In these models, cultural traits, however they arise, spread as fitness-enhancers, at least locally considered. Agriculture, cooking, clothing, the wheel, writing, bow and arrow--all these cultural innovations are plausibly viewed as improvements that need not arise from gene mutation and recombination and need not be transmitted genetically. They are, one might say, infectious phenotypic features. These features, it is presumed, pass some sort of quality-control test administered by the agents themselves. They are chosen by evolved organisms and put to use, and if they didn't "pay for themselves" in a fitness boost (or at least an apparent fitness boost, myopically considered) they would soon die out, just like genetically transmitted phenotypic variations. The idea is that if innovations are randomly distributed around neutrality, the pernicious innovations will hasten the demise (or mating failure) of those who adopt them, and the enhancements will do the opposite, and over the not very long haul the enhancers will proliferate. This vision allows the possibility, as relatively rare outliers, of mistakes: either good tricks abandoned by mistake or pernicious tricks persisted in thanks to some local illusion.

More ambitious models (Feldman and Cavalli-Sforza, 1981; Boyd and Richerson, 1985) then address the opportunities for co-evolution, for interaction between the items that come to be present or dominant in the cultural marketplace and the genetically transmitted phenotypes of those transmitting and preserving those items. Clothes do make the man, at least to the extent of diverting selection pressures for weather-hardiness, so the cultural transmission of clothing sends ripples through the evolution of human physiology. Similarly, new practices of food gathering and preparation can reflexively change the fitness landscape for digestive prowess. Lactose tolerance in adults descended from people who herd dairy animals is a well-studied parade case. These models are Darwinian in a more than minimal sense because they extend the perspective of population genetics, the replicator dynamics, to these non-

genetically transmitted phenotypic features, exploring the effects of horizontal and oblique transmission, for instance. But they also maintain the basic economic presupposition of agent rationality: cultural traits are *adopted* because they are *deemed worth having*, because they are supposed, rightly or wrongly, to contribute somehow to the achievement on one's more ulterior ends, whatever they be.

When a rational 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. Here things begin to get messy. In nature, genes are the ultimate units of "self"-interest. That is to say, adaptations in plants and animals (and simpler life forms) are, by definition, features that further the *summum bonum* of gene replication, directly or indirectly. Are cultural innovations adaptations in this narrow, genetic sense? It is obvious that many cultural features are deemed by the populace to be beneficial, functional, adaptive, useful, life-enhancing, enabling. It is less clear that these esteemed features play a discernible role in enhancing *genetic fitness*, as contrasted with, say, human happiness, the pursuit of which is curiously orthogonal to genetic fitness. One of the striking trends in human evolution, going back thousands of years, is the gradual diminution in the proportion of human effort devoted in any clearly discernible way to the achievement of the fundamental goals we share with animals: avoiding pain, hunger, and predation, and seeking comfort, security, and mating opportunities. Even if the peculiar human *desiderata* of prestige, power, wealth, beautiful surroundings, recreation, music, toys, . . . have discernible instrumental rationales (improving one's profile in the contest for mates, enlarging one's harem, one's territory, one's margin of error), they have more or less detached themselves from these inaugural foundations and become ends in themselves. The young man bought the guitar in order to attract young women, but now he has become a guitarist who would rather make music than love.

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. (p.342).

As they put it,

Control is delegated to a system of poorly understood internal drives and rewards that direct the activity of the individual, . . . 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. (p364.)

Feldman and Cavalli-Sforza thus adopt the default assumption, at least in the Western world, and

especially among economists, of treating 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*. 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, the flourishing of mid-twentieth-century acoustic guitar finger-picking techniques . . . anything human ingenuity can define and become attached to, making its welfare definitive of the decision-maker's "happiness".

It is not obvious that any other organism strives for its own happiness or anything like it. If human happiness is our *summum bonum* (or at least a *bonum* against which we do in fact often attempt to measure costs and benefits), how did it arise? It is here that the prospect of a still more radically Darwinian theory of cultural evolution becomes attractive. Could the unique varieties of human evaluation that are so distantly and indirectly anchored to any plausible litmus test of genetic fitness be accounted for by supposing that human beings have evolved into a condition where they have become the vectors, the hosts, of a new order of symbionts, competing cultural replicators whose *own* fitness, defined in standard Darwinian terms of relative replicative success, has constituted a new sort of entity? An enculturated human being, according to this proposal, is as important a novelty on the evolutionary front as the eukaryotic cell was at its debut: a unification of distinct replicators into a synthetic organization with a displaced goal or *summum bonum*, no longer just an *organism* bent on self-replication, but a *person*, bent on furthering the particular goals and ideals with which that person identifies. Has our guitarist unwittingly become part of a guitar's way of making another guitar? This is a tantalizingly attractive idea, but for such a perspective to anchor itself firmly in evolutionary theory, we must take seriously Dawkins' concept of a meme, and there are reasons to doubt that we should.

2. Cultural Replicators: A Central or Peripheral Phenomenon?

In some neighborhoods, ball bearings outnumber rabbit turds; in other neighborhoods, this imbalance is reversed, and in yet other neighborhoods, no entities of either variety are to be found. These differential production patterns change over time, and there are reasons for them, but they are not, in the main, Darwinian reasons. Not all production is replication, and not all distribution is emigration. Variety and similarity is also found among cultural items, and the question is: are any (or many) of the reasons for patterns in changing "populations" of cultural items Darwinian?

Dan Sperber (this volume) notes that the dictionary definition of "memes" is too bland to be of much interest: "an element of culture that may be considered to be passed on by non-genetic means," while a more radical definition, more faithful to Dawkins' arresting proposal, "cultural replicators propagated through imitation," is far from obvious. Indeed, it is in need of defense against two objections. The

"simplest and most serious objection" is that the copying of cultural items is in general too low-fidelity to permit natural selection to get a purchase. Compare memes to viruses. Viruses travel light, and carry no copy-machines, so they reproduce by entering cells and tricking the cell's proprietary copy-machines into making spurious copies of them instead of copying their usual and proper fare, the cell's own DNA. If memes are like viruses, as Dawkins and other would-be memeticists have claimed, it is because they reproduce by exploiting the copy-machinery resident in the brains of human beings. But how well does this parallel hold? How good is that machinery? Not good enough, it seems. We human beings are actually rather bad at the sorts of "mindless" copying that cells excel at.

Following Williams (1966), Sperber notes that although higher selection biases can tolerate lower fidelity, it still appears that "mutation" rates among memes would be so great that any description of the emerging patterns in terms of descent with modification, as Darwin put it, would be lip-service only. "For memetics to be a reasonable research program, it should be the case that copying, and differential success in causing the multiplication of copies, overwhelmingly plays the major role in shaping all or at least most of the contents of culture." If we are not inveterate and talented copyists, we will make poor hosts for our cultural viruses, and Darwinian descent with (relatively rare) modification will seldom occur. We will need to look elsewhere to explain the patterns of culture.

But is it so clear that our copying is too low-fidelity to work? Dawkins (1999) has responded to this objection with his example of the origami model of a Chinese junk, which people learn to make by following a canonical set of simple "self-normalising" instructions, but Sperber finds this misleading, since "the normalisation of the instructions results precisely from the fact that something other than copying is taking place." Sperber lays down three conditions for "true replication":

For B to be a replication of A,

- (1) B must be caused by A (together with background conditions).
- (2) B must be similar in relevant respects to A
- (3) The process that generates B must obtain the information that makes B similar to A from A.

It is condition (3), Sperber claims, that is seldom met by cultural transmissions. Infectious laughter is his excellent example of a transmission event that meets (1) and (2) and fails (3), and he extends his analysis of this case by the fanciful example of the ten sound-recorders that trigger each other, but whose productions, in one case, do not consist of replications, but rather of *recognitions*, followed by *reproductions*. Triggered reproduction of this sort is distinct from copying or replication in the one way that matters for Darwin, according to Sperber: it does not slavishly copy the original; instead it is inspired by the original to make another of the same sort-but *without* any systematic attempt or disposition to reproduce any idiosyncracies of the original. It normalizes its production to an independent ideal, discarding or not even noticing any mutations, good or bad, in the original.

Sperber illustrates this point with another fine example, the contrast between the nonsense scribble and the five-pointed star. The nonsense scribble would degenerate quickly in any series of attempted replications because people are not good copiers of such productions, while the five-pointed star would be "copied" with high fidelity, just as Dawkins says, but, Sperber maintains, the succession of stars would not really be copies of their predecessors, since the "copyists" would normalize to the recipe for the drawing procedure, ignoring the details of the individual productions. But is Sperber looking at the right grain level? Dawkins' point is that a finite repertory of such triggered reproductions is not just a good trick for human beings who want to heighten their transmission fidelity; it is a Good Trick already discovered, several billion years ago, by natural selection. Sperber distinguishes copying from "merely triggering the production of a similar effect", but a repertory of such triggers, called an alphabet, is just what makes high fidelity copying possible, both in cells and in human culture.

Suppose Tommy writes the letters "SePERaTE" on the blackboard, and Billy "copies" it by writing "seperate." Is this copying or triggered reproduction? The normalization to all lower-case letters shows that Billy is not slavishly copying Tommy's chalk-marks but rather being triggered to execute a series of canonical, normalized acts: *make an 's'*; *make an 'e'*, etc. It is thanks to these letter-norms that Billy can "copy" Tommy's word at all. And he *does* copy Tommy's spelling error, unlike Molly, who "copies" Tommy by writing "separate," responding to a higher norm, at the level of word spelling. Sally then goes a step higher, "copying" the phrase "separate butt equal"-all words in good standing in the dictionary-as "separate but equal," responding to a recognized norm at the phrase level. Can we go higher? Of course. Anybody who, when "copying" the line in the recipe "Separate three eggs and beat the yolks until they form stiff white cones," would replace "yolks" with "whites" knows enough about cooking to recognize the error and correct it. Above spelling and syntactic norms are a host of semantic norms as well.

DNA has an alphabet, the famous ACGT, and words, the three-letter codons that "spell" the twenty amino acids. In fact, the high-fidelity of genetic transmission depends on the sub-cellular machinery being triggered to "recognize" and "re-produce" a small repertoire of types, whose idiosyncracies, if any, are ignored, not slavishly copied: "*make a cytosine*", "*make a guanine*", etc. There are error-correcting enzymes as well, but they don't ascend (so far as we know!) above the level of a spell-checker, correcting "typos" by brute template-matching against the original.⁽³⁾

Does the human capacity (and irresistible disposition) to respond to higher, semantic norms-our capacity to correct not only typos but what hackers call *thinkos*-rule out cultural transmission as a candidate for natural selection? Sperber seems to think it does. "Contrary to what Dawkins writes," he claims, "the instructions are not 'self-normalizing'. It is the process of attribution of intentions that normalizes the implicit instructions that participants infer from what they observe." [ms p8] Sperber is partly right: the attribution of intentions is the key difference between this sort of human transmission and genetic replication. The point comes out even more clearly if we mutate Sperber's example slightly, adding a point to his star: Consider the fact that there are two distinct recipes (and many other less obvious ones) for making a regular six-pointed star:

(A) make a regular hexagon and put equilateral triangle points on each side.

(B) make an equilateral triangle and the superimpose on it another one, upside-down.

A series of six-pointed-star "replications" might be accomplished by a random alternation between these two recipes with no loss of fidelity. Which recipe did various individual copyists follow? It wouldn't matter, because what is being copied is not the recipe but the result *understood as an intended object having certain features.*⁽⁴⁾

Sperber thinks that this reliance on attribution of intentions on the part of the copyists disqualifies cultural transmission as a Darwinian process of natural selection. He supposes that this invocation of intelligent, semantically-sensitive, intention-attributing agents in the purported replication process flies in the face of a fundamental requirement of Darwinian processes: mindless, purposeless mechanicity. He is almost right. To see the force of this interesting objection, imagine a creationist variant on standard neo-Darwinian genetic evolution. It postulates that God watches over each moment of DNA-replication, and whenever He sees some copying that is "wrong" (relative to God's great plan), he undoes it. Thus when He chooses, he lets mutations flourish, and when He does not, those mutations get corrected by a gentle miraculous nudge of the error-correcting enzymes. Here Intelligence is playing a guiding role in evolution-just the sort of role that orthodox (devout, "fundamentalist") Darwinians abjure. As Richard Powers (199x) has observed, "Natural selection edits with an eye only toward what the message says, not to what it means." Clever human beings, in contrast, edit with an eye toward meaning. If such clever editors are inserted into the process of cultural transmission and revision, what would be left of a Darwinian theory of culture?

This worry ignores the fact that *Homo sapiens* is not itself a miracle, a skyhook, but something that has evolved by non-miraculous natural selection; its capacity to respond to semantic norms is itself something that has evolved under a regime that could not respond to semantic norms. Before there could be eyes, good for distal perception, there had to be mere photo-sensitive responders to proximal stimulation, out of which eyes could gradually be built. Before there could be minds, good for semantic discrimination, there had to be copying-machinery that could only discriminate alphabet letters. Put otherwise, DNA error-correcting enzymes have always responded to semantic norms, but just *local or proximal* semantic norms-*make a "G"*-as contrasted to more *distal* semantic norms: *make a codon for asperagine* or *make some lysosyme* or *make a protein that blocks seratonin uptake*, or even *make something that will fight off infection*.

Why shouldn't evolution go right on working once the copying machinery graduates to less myopic norms? Even our lowest-level mindless copying avails itself of correction to a norm; is there a "highest permissible" level of normalization in any Darwinian process? Darwin (and Fisher, and Williams, and others) saw the need for a sufficiently "strong principle of inheritance" to keep evolution going, but nothing has been said about how that fidelity is to be maintained, mechanically. Let there be copyists that take themselves to be responding to semantic norms; there will still be a suitably long-distance evolutionary perspective from which their copying efforts, for all their editorial work, will appear myopic and unwitting, oblivious to-and hence unresponsive to-the larger scale pattern of differential replication that ensures that a Darwinian process is occurring.

In "Pierre Menard, Author of the *Quixote*," (1962) Jorge Luis Borges tells the fanciful tale of a literary theorist who sets out to *compose* (not copy, not write from memory) Cervantes' great work anew in the 20th century. This will be an act of bizarre self-control, since Menard is a Cervantes scholar who no doubt has at least large portions of the text of *Don Quixote* committed to memory, but Menard is determined to bracket that memory and *create*, with his own authorial intentions, all of Cervantes's sentences anew, like an experienced wheelwright setting out to *re-invent* the wheel! He succeeds (though how can he tell?), and Borges tells us: "Cervantes' text and Menard's are verbally identical, but the second is almost infinitely richer." (p42) In one sense, Menard did not *copy* or *memorize* Cervantes' text, but in another sense, he did, in spite of his virtuoso self-control, his obsessive act of re-creation. He did, because, as Sperber puts it, "(3) The process that generates B must obtain the information that makes B similar to A from A." and surely Menard's prior study of Cervantes' text is an essential part of the scholarship that permits him to "compose" *Don Quixote* anew. Of course Menard has used a lot of other information as well; the surplus is presumably what permits him, unlike an ordinary reader, to claim to have re-composed, not written down from brute memory, the work. But so what? According to Borges, the texts are "verbally identical" so high-fidelity reproduction has occurred. Imagine a world in which Menards abound, devoting their lives to the re-composition of their favorite works. The transmission of texts will proceed just fine in such a world-as fine as if photocopy machines were the underlying machinery.

In fact, of course, a pastel version of that fantasy is just what has happened in the transmission of ancient texts in our world, for seldom if ever have the scribes taking dictation been entirely uncomprehending of the words they were dutifully "copying," and so they have willy-nilly "corrected" whatever they heard in the process of transcribing it. Their corrections have been governed by several levels of norms: orthographic/lexical, syntactic, and finally semantical. The imaginary Menard can be conceived to have "transcribed" the entire poem of Cervantes *modulo* the "semantic norm" of the whole text. Most of us lack that highly sophisticated norm; we tend to fall back on our sense of the gist of such a narrative, or when all else fails, rote memory or parroting (but even "parroting" is not like a parrot's parroting-unless it is, as it very seldom is, a matter of reiterating formulae in a language we don't understand).

When Sperber notes that in cultural transmission "the information provided by the stimulus is complemented with information already in the system" he is right, but the same is true of DNA replication. The main difference, so far as I can see, is that unlike DNA replication, human cultural replication is accomplished by processes of highly variable semantic depth, responding to perceived (and mis-perceived) "copying" errors relative to norms at many levels. The alphabets of written languages provide us with the most vivid and best understood system of such norms of replication, but the phenomenon of semantic norms is not directly tied to language. Musical notation relies on the staff to digitalize the roughly inked spots, so that a musician can see at a glance that a chord is A-C#-E-G even though the A is written almost twice as far beneath the staff as it is "supposed" to be. A sketch of a new sort of axle for a wagon need not make the wheels exactly round; the user of the sketch will recognize those irregular closed curves as representations of wheels, which are to be round, of course. As we move through our various apprenticeships in life, we learn to perceive new families of categories--new alphabets, in an extended sense--from which to construct high-fidelity copies. Only a skilled potter can see at a glance what another potter is doing and copy it, or teach it to others.

