Gil-White, F. 2008. "Let the meme be (a meme): Insisting too much on the genetic analogy will turn it into a straightjacket," in *Culture, Nature, Memes.* Edited by T. Botz-Bornstein, pp. 158-190. Newcastle upon Tyne: Cambridge Scholars.

Let the meme be (a meme)

Insisting too much on the genetic analogy will turn it into a straightjacket

Francisco J. Gil-White

> Given an incredibly simplistic notion of genes, memes are not in the least like genes. . .One problem with interdisciplinary work is that any one worker is likely to know much more about one area than any of the others. Geneticists know much more about the complexities of genetics than of social groups. Conversely, anthropologists and sociologists tent to be well-versed in the details of social groups. To them genetics looks pretty simple.—Hull (2000:4)
>
> Many of the claims made about memes could be false because the analogy to genes has not proven productive.—Aunger (2000:8)

Introduction

Following Dawkins (1989[1975]), many now refer to units of cultural transmission and evolution as 'memes' (the 'genes' of culture), they claim that any Darwinian process requires 'replicators' (because genes are replicators), and they adopt a 'meme's eye view,' which, by analogy to the 'gene's eye view,' implies that memes must be 'selfish.' It can be quite productive to borrow terms by analogy and reproduce the yardsticks of a sister field, already developed, but this can be taken too far and cause misunderstanding. The question is: Should we demand that 'memes' be *exactly* like genes if we are to apply Darwinian tools of analysis to culture? I will argue that this would sell Darwinism short.

The formal similarities between genes and memes suggest that cultural transmission processes are ripe for Darwinian analysis, and a vigorous debate has emerged over how to think about memes (see Aunger 2000). There are many dangers here. Especially in the social sciences, careers are often boosted by getting particular definitions adopted, which results in tremendous indiscipline in the use of technical terms of analysis (Gil-White 2006). And the prize in this case is rather large because the *term* 'meme' is already on everybody's lips. If definitions were advanced only with conceptual progress in mind, this would be fine. But the various protagonists in this field must be at least vaguely aware that the contest is memetic, yielding a tendency to produce 'catchy' definitions that 'sell well' at the expense of conceptual advance and scientific utility. Although the definition of meme as 'replicator' is catchy, what ultimately matters is whether it is scientifically useful.

A. What accounts for *genetic* cumulative evolution?

Darwinian systems involve simple and blind algorithmic processes that nevertheless produce gradual accumulation of (sometimes very complex) adaptive—that is, purposeful—design. They have three main requirements: information must be able to leave descendant copies (*inheritance*), new information should be routinely generated by some process (*mutation*), and there should be forces responsible for causing some items of information to leave more descendants than others (*selection*).

Genes satisfy all three. They are inherited through reproduction; new genes are routinely created because of occasional copying mistakes, or 'mutations', during DNA duplication; and a gene, through its effect on its carriers, affects the probability that it will increase in number. Thanks to selection and inheritance, when a particular gene causes increased reproductive success, more copies of it are passed on, and its relative frequency in the population increases (absent frequency dependent effects, eventually the whole population will have it). Thanks to mutation, new alternative genes get generated which occasionally amount to improvements, allowing the population to continue to evolve.

Here is an important general principle: the space of maladaptive designs is vast relative to the space of adaptive ones, so *random* changes to any current design are unlikely to cause adaptive improvements. For example, this essay is an example of an adaptive design—it has a point. Imagine that a monkey types a character at random as I am editing it. Will it improve the way in which my essay gets across the point? Without vanity, I can say that the chances are exceedingly low. A random typo is unlikely to yield English, let alone better English. Something else is obvious: should the monkey press a key which launched a program to rearrange at random *all* of the letters in my essay, then he would be infinitely less likely to improve it—slim as his chances were anyway. The bigger the mistake in my essay, the higher the chance that I will throw it away, so the 'survival time' of *small* random mistakes is better than it is for larger ones.

Genetic mutations, it appears, are random with respect to reproductive success: as if a monkey typed at random in the DNA. The analogy holds: if the monkey types at random in several thousand genetic loci, causing massive changes, natural selection will be more likely to 'throw away' her 'essay.' Random novelties are usually detrimental to the point of an adaptive design—here, reproductive success. And sudden and massive changes will tend to be *lethal*—not merely detrimental—to reproductive success. So if a random change is to be so lucky, after a shift in the direction of selection, to become an adaptive feature, it must first of all not get immediately deleted. We should expect genetic evolutionary change, therefore, to consist of small, incremental changes.

This is why genetic evolution is *cumulative.*

But mutations must also be *infrequent*, because otherwise designs will not be sufficiently *stable*. Suppose the offspring of A's are mostly non-A's. Even if A reproduces better than

its competitors B and C, this cannot have an evolutionary consequence because the information responsible for A's reproductive prowess is almost always lost after reproduction. On the contrary, if an A typically begets another A, then A's higher reproductive success will soon make everybody in the population an A (absent frequency-dependent effects). Later, when a rare mutation results in a slight improvement to 'A design'—let us call the new design A°—these A° mutants will outreproduce mere A's and the population changes again (but only slightly).

Cumulative genetic adaptations, then, are possible because (1) genetic mutations typically introduce incremental rather than massive changes, and (2) the rate at which genetic mutations occur is low.

B. How similar to genes are memes?

Memes certainly have the properties of inheritance, mutation, and selection. We constantly acquire and learn things from each other through social interaction, so in a broad sense at least it makes sense to say that the acquired information I possess can create a 'descendant copy' in you (*inheritance*). People can make mistakes when acquiring information, and can also have stupid or bright novel ideas, which leads to new items of information (*mutation*). And some ideas are more popular than others, so they are copied more, stored longer, and rebroadcast more often, which in turn means they leave more descendants than competing ideas (*selection*). What makes some ideas more 'popular' than others are the properties of human social-learning psychology. This is not the only force acting to favor certain memes over others, but it is a very important one and I shall restrict myself to it here.

So much for intuitively stated formal similarities. The devil, as usual, lurks in the details. To many critics, the dangerous phrase above is "in a broad sense…acquired information…can create a 'descendant copy.'" How broad? How similar must ancestor and descendant memes be?

Some assert that selectionist approaches cannot work because memes are not true replicators, making cumulative evolution impossible (e.g. Sperber 1996; Boyer 1994). Others, however, have not considered this a problem and proceeded to build Darwinian selectionist models that in their fundamental assumptions are quite similar to those used in evolutionary genetics, but adapted for the idiosyncrasies of culture (e.g. Boyd & Richerson 1985; Lumsden & Wilson 1981; Cavalli Sforza & Feldman 1981; Castro & Toro 2002; for a review, see Feldman & Laland 1996). As Laland & Odling Smee (2000:121) put it: "For us, the pertinent question is not whether memes exist [as such]… but whether they are a useful theoretical expedient." Their critics, however, will counter that such models do not help us explain human cultural processes because the units employed are nothing like what exists in real-life cultural transmission.

It is true that memes typically do not replicate. Now we need to decide whether this matters. Before we do this, however, I must clear up some terminological confusion, or I fear I will not be understood.

C. Making sense of our terms

What did Richard Dawkins originally mean by 'replicator'? Not surprisingly, something that produces "identical replicas" (Dawkins 1989:16). His way of putting it is circular, but consulting any dictionary we can see what he meant: a 'replica' is "an exact reproduction," that is, "a copy exact in all details."[1] Of course, Dawkins emphasized the importance of copying mistakes because without them no Darwinian evolution is possible. But he was hardly arguing for an entity that makes a mistake every time it copies itself. For him 'replicators' are "astonishingly faithful" copiers which only "occasionally make mistakes" (Dawkins 1989:16-17).

Dawkins originally argued that the very reason genes set in motion Darwinian evolutionary processes is that they are replicators. He carefully rehearses this point for his audience by asking "What, after all, is so special about genes?" to which his answer is "that they are replicators." Only after going through this rehearsal does Dawkins introduce the idea of "a new kind of replicator [that] has recently emerged…" (Dawkins 1989:191-192). *Enter the 'meme.'* The argument, therefore, is obviously that cultural transmission is Darwinian *because* memes—too—are supposedly replicators. Any doubts about Dawkins' position on this point are removed when he observes that apparently "meme transmission is subject to continuous mutation, and also to blending," which he immediately identifies as a potential *problem*, for it appears to contradict that memes produce exact reproductions or 'replicas.' He suggests, however, that "it is possible that this appearance… is illusory, and that the analogy with genes does not break down" (Dawkins 1989:194-195). At least when he wrote *The Selfish Gene*, then, Dawkins held that a Darwinian process absolutely *needs* replicators analogous to genes.

If this were a rigorously stated hypothesis, it wouldn't matter if it were wrong, for it would suffice to investigate the properties of memes and then we could discard it. But in fact some self-conscious followers of Dawkins have taken his *word* 'replicator' without the concept it stands for. For example, Susan Blackmore (2000:25) first argues for a *non*-replicating definition of 'meme' as follows: "As long as we accept that people do, in fact, imitate each other, and that information of some kind [!] is passed on when they do, then, by definition, memes exist." But in the next page she says that "memes are replicators." Substituting the second statement into the first what she is saying is that whenever any information is passed on, no matter how loosely, then—by definition!—we have a *replicator*.

This is a new definition. My dictionary—and Richard Dawkins (1989), from whom Blackmore claims to get the term 'replicator'—both insist that 'replication' takes place when perfect copies are produced, not when "information of *some kind* is passed on…" Blackmore is keeping the word 'replicator' but chucking its meaning. And yet Dawkins' requirement, which Blackmore pretends to be applying, was that the units of a Darwinian process must *be* replicators—not that they must be *called* 'replicators.' To act as if the

[1] Merriam Webster's Collegiate Dictionary.

mere use of the term 'replicator' will make the inheritance unit so-dubbed Darwinian is not science. It is magic. And Blackmore is not alone: Daniel Dennet (1995) does the same, and Robert Aunger (2002:3) likewise defines replication as "the recurrence of… features," eliminating all emphasis on exact reproduction.

It seems that Dawkins has followed his followers (an interesting memetic process!). In his introduction to Blackmore's 1999 book, he now dispenses with his earlier worries, and declares that memes are definitely subject to Darwinian analysis. Either (1) Dawkins sticks to his requirement that genuine units of Darwinian analysis be bona-fide replicators, and he convinces himself that such are memes; or (2) Dawkins sticks to his belief that memes do not replicate, but convinces himself that memes can be subject to Darwinian analysis *despite* this failing. Either avenue would produce a logically consistent argument. However, the result has been a mix and a match between the two options that defies logic. Dawkins' examples in his introduction to Blackmore's (1999) book are explicitly and self-consciously about things that do not replicate, and he proceeds to explain how this does not matter. The explicit conclusion should be: "Lo! Replication was not necessary for Darwinian processes." But his explicit conclusion—after Dennet and Blackmore—in fact is: "Lo! Memes are replicators, after all."

What happened? Just this: the world was turned upside down. We started with 'replication' as the proposed requirement for the possibility of Darwinian analyses, and we ended up with a 'principle' that goes the other way: we ask first whether a particular unit can be the object of Darwinian analyses and, if it can, then we baptize it 'replicator' with nary a thought given to its copying fidelity! The *meaning* of 'replicator' has thus been transformed from 'unit that copies itself perfectly' (i.e. produces replicas) to 'unit of Darwinian analysis' (whatever its copying fidelity). But why do this? It is exceedingly confusing because 'replication' is, after all, a common English word whose meaning has everything to do with copying fidelity, and this is precisely why Dawkins originally recruited its meaning. The reason for torturing language this way seems to be that the marketing of 'replicator' was so successful—so 'catchy'—that Dennett, Blackmore, and Dawkins cannot get themselves to abandon the *word*, even as they make it stand for an entirely new concept. They want to win the memetic contest.

Meanings have been turned upside down. We started with replication, defined as near-perfect copying fidelity, as the requirement for Darwinian processes. But these writers now ask first whether a unit is Darwinian, and if they find that it is, they call it a 'replicator', whatever its copying fidelity. As a result of this extra-scientific semantic flip, the view that replication is a requirement of Darwinian processes has become entrenched. The lamentable consequence is that some critics have found it easy to dismiss the very idea of a Darwinian approach to culture merely by pointing out that memes are not, in fact, replicators (e.g. Sperber 1996, Atran 2001). They are sticking closely to the definition of 'replication' as 'perfect copying,' and this much is good (why butcher the language?). And memes indeed don't replicate. But these critics are being 'more Catholic than the Pope' by ferociously upholding the 'gene-analogy standard' and doggedly adhering to Dawkins' *original* argument that Darwinian processes require replication. As I will argue here, particulate replication along genetic lines is not necessary for

cumulative adaptations through selection in cultural processes (cf. Boyd & Richerson 1985:75, 2000:153-158), and this is therefore not the litmus test that such critics are looking for. Replication is a red herring.

The 'selfish meme,' like its ancestor the 'selfish gene,' is another catchy idea (another good meme!), and again a legacy of Dawkins'. The ancestor concept was developed to answer the question *cui bono?* Who benefits from Darwinian organic selection? It is the gene, because bodies live once and never return, but genes are potentially eternal. Organisms are merely the 'survival machines' that genes have built to get themselves from one generation to the next. The 'gene's eye view' is the perspective that results from accepting the 'selfish gene' argument, and it gives us the correct answer to organic evolutionary problems by stating them in terms of the gene's proliferation 'interests.'

By analogy, Dawkins' argues that in cultural evolution it is the *meme*, and not the human vehicle that benefits. In Dawkins' (1983:109) own words, a meme is "a unit of cultural inheritance… naturally selected by virtue of its… consequences *on its own* [my emphasis] survival and replication." A meme thus succeeds if it is "*advantageous to itself*" (Dawkins 1989:200; original emphasis). Cultural selection is understood to be the continuous editing of a meme's content until it becomes optimally designed for colonizing human brains. Dennett (1995) and Blackmore (1999, 2000) have developed this idea further, giving us a picture in which "We humans… have become just the physical 'hosts' needed for the memes to get around. This is how the world looks from a 'meme's eye view'" (Blackmore 1999:8). Aunger (2002:12-13) states that "it is only when information *replicates* that an additional causal force becomes involved. This is the very essence of the meme hypothesis… there is an information-bearing replicator underlying communication… a puppeteer pulling invisible strings…This puppeteer is the information packet itself, evolved to manipulate its carriers for its own ends." This statement brings together all of Dawkins' influences.

I think only some rather specialized kinds of memes are really 'selfish' in the sense these authors intend. But, again, this does nothing to wreck the validity of understanding culture as a Darwinian process—it merely indicts importing willy-nilly the 'gene's eye view' from biology (where it is perfectly sound) to culture. I will argue here that reducing all cultural transmission to 'selfish memes' requires that we ignore much social-learning cognition and miss most of the picture.

It should be obvious this far that I feel no compunction to accept the definition of 'meme' as *selfish replicator*. Neither do several others working in this field (who expressed their views in Aunger 2000). Most of us seem to accept the broad Oxford English Dictionary's definition, which says: 'an element of culture that may be considered to be passed down by non-genetic means.'[1] Thus, I will treat 'selfish replicator' as a *hypothesis* about what the stuff that gets transmitted through non-genetic means is like. The relevant questions, then, are: (1) does this stuff look like a selfish replicator?; (2) If not, does this really make Darwinian analyses of culture impossible? Related questions are: (3) if they don't replicate, is it impossible to find the boundaries of memes?; and (4) can we import from biology, willy-nilly, the 'selfish gene' idea? I will answer "no" to each of these questions.

Nevertheless, I will still call what is transmitted culturally a 'meme,' and so—I will bet my house—will everybody else. The *term* 'meme' has already been selected for, so rather than forcing its meaning to coincide with a particular hypothesis about cultural transmission, let us do some science.

Do memes mutate too much?

Genes *replicate* because they almost never make copying mistakes during duplication. As mentioned earlier, genetic replication produces good stability for the overall design template, which in turn allows for cumulative adaptations to emerge when small, incremental changes are added at the margins. But it is important to see that there is nothing absolute about the acceptable rate of mutation. Rather, this is always relative to the strength of selection. For example, even if there is a moderate rather than low rate of mutation, cumulative evolution will still happen if the selective process is so strong that it culls suboptimal variants fast enough to make the favored design stable at the populational level by eliminating the continuously popping mutant variants. G.C. Williams (1966) made this principle famous in his definition of an 'evolutionary gene,' which is "any hereditary information for which there is a favorable or unfavorable selection bias equal to several or many times its rate of endogenous change."

If we apply these concepts willy-nilly to memes, then we will produce a definition like the one given by Wilkins (1998:8): "A meme is the least unit of sociocultural information relative to a selection process that has favorable or unfavorable selection bias that exceeds its endogenous tendency to change." If we accept this definition of the meme, it follows that the idea of Darwinian analyses of culture can be challenged by arguing that the meme's "endogenous tendency to change"—its rate of mutation—is so high that it creates a ceiling effect. Dan Sperber, for example, argues that "recall is not storage in reverse, and comprehension is not expression in reverse. Memory and communication transform information" (Sperber (1996:31). Does anybody ever retell a story *exactly*? No: memes mutate *every time they are transmitted*, and no selection pressure can keep up with a rate of mutation = 100%.

> In the case of genes, a typical rate of mutation might be one mutation per million replications. With such low rates of mutation, even a very small selection bias is enough to have, with time, major cumulative effects. If, on the other hand, in the case of culture there may be, as Dawkins [1976] acknowledges, 'a certain "mutational" element in every copying event,' then the very possibility of cumulative effects of selection is open to question.—Sperber (1996:102-103)

So, Sperber is accepting the move to assume (1) that 'replicators' are the things to look for; (2) that from Williams definition of an evolutionary gene we can abstract a *universal* definition of a Darwinian replicator and generate the definition of an evolutionary meme; and (3) that Darwinian analyses will apply in cultural transmission *only if* the stuff that gets transmitted satisfies this definition. In fact, Sperber eagerly forces the issue by ruling that any other conceptualization of 'the meme' is trivial (Sperber 2000:163). Summarized, his stance is that cumulative adaptations through cultural selection are

possible *only if* we can find bona-fide cultural replicators, but since memes mutate in *every single act of transmission*, no selection pressure can be high enough to make them replicators (Atran 2001 echoes this argument). A selectionist perspective in culture is useless; we must understand how cognitive processes of information storage and retrieval cause mutations in particular and systematic directions. With this information, we can build (orthomemetic?) models of 'directed mutation' rather than selectionist models of cumulative change (Sperber 1996:52-53, 82-83; 110-112).

There is irony here. Hull (2000:47) quotes Wilkins' definition of an evolutionary meme approvingly as a starting point for a science of memetics that he optimistically believes to be *possible*, although he fully expects "howls of derision" to come from unreasonable critics who will accuse this definition of not being sufficiently "operational." Something very different has already happened, however. A prominent critic of selectionist approaches to culture—Dan Sperber—has *eagerly embraced* that very definition in order to explain why selectionist approaches to culture are supposedly *impossible*.

It would seem as though either Hull or Sperber must be wrong, for they agree on how to define units of cultural processes that would be legitimately Darwinian, but they reach exactly opposite conclusions as to whether human culture passes or fails the test. However, I believe they are both mistaken because they are sparring on the wrong battlefield. The standard chosen, rather than enlighten, blinds us to the general requirements for a Darwinian system by insisting narrowly on the terms of one particular solution to them—the genetic one—as if this were the only possibility.

I shall accept Sperber's point that the mutation rate for memes = 1: they mutate in every act of transmission. And I will agree, too, that often they are systematically biased. But neither point automatically disqualifies the possibility of cumulative memetic evolution through Darwinian selection. What matters is how big the mutations are, and how strongly biased in particular directions, as we shall see.

'Replication' is a red herring

Though Sperber's argument may seem intuitively appealing, I think it is specious. Near-perfect copying fidelity is certainly important in *genetic* selection, but it is not *the* requirement for *any and all* Darwinian systems. So long as a high rate of mutation is not the meme's only distinction, its other idiosyncrasies in principle could make it possible for regularly imperfect—or even invariably imperfect—meme-copying to support the emergence of cumulative adaptations. This will be my argument.

I shall make the case with a toy example. First, a few preliminaries. In genetics, a 'locus' is the physical location of a 'gene' on a chromosome. This is where the information 'for something' can be found. If we are talking about, say, the 'eye-color' locus, then the gene found there may be the 'brown-eye' gene, or the 'blue-eye' gene, and so forth. What is the analogue in memetic transmission? For example, imagine something like, say, a tennis-serve 'locus.' Whatever is in your tennis-serve locus will produce a behavior when it is your turn to begin a new tennis point. There is in principle a vast number of different

behaviors that people could store at the tennis serve locus (just as there are many different sequences of nucleotides that may be stored at the chromosomal eye-color locus). Waving hello to your mom, or baking a bread, would be ruled illegal by the judges, but in principle this does not prevent you from storing such information at that locus (just as a random and useless sequence of nucleotides could, in principle, be stored at the eye-color locus).

It hardly matters that the tennis-serve locus may not be physically located in the same piece of brain for every individual. To insist on this is to push the genetic analogy to an absurd extreme where it begins to straight-jacket thought rather than inspire Darwinian insights. The relevant and crucial similarity is *functional*, not physical: if individuals recognize that an item of information becomes relevant when, in a game of tennis, a new point is beginning, then the 'cultural locus' has all the requisite *functional* similarity to the genetic locus that we need. In cognitive terms, the cultural 'locus' is a *tag plus retrieval function*—it is a matter of categorization rather than physical location in the brain. The information retrieved at the start of a new tennis point is that which I tag as 'tennis serve.' Waving to my mom or baking a cake have not been tagged this way (even though, in principle they *could* be), and, since they have not been, they do not compete to 'occupy' my tennis serve 'locus.' The true *alleles* of my current serve, therefore, are *other* behaviors which I also tag as 'tennis serves' because some individuals in the population perform them in the context of beginning a point in a tennis match. I may choose to acquire one of these later on, and in so doing will replace my current serve.

These obvious functional similarities readily dismiss the criticism that, because memes do not have the same kind of physical reality as genes, selectionist approaches to culture are a nonstarter. We are not talking here of the duplication of exact neuronal structures analogous to the duplication of exact nucleotide sequences in DNA, but we *are* speaking of the duplication of a certain behavior, understood to belong in a certain context, and in competition with other behaviors also understood to be candidates for the same context. The lack of similarity in the *material* basis of genes and memes is not a problem.

D. The right mix of stability and variation

To see whether a meme's inability to replicate makes cumulative adaptations impossible in culture, we must examine the full spectrum of theoretical possibilities.

Suppose that in our population, Bob's serve is the most attractive, and seeing it performed gets people excited to make changes in their own tennis-serve loci. There is a continuum of different things that could happen, bounded by two extremes. At one extreme—replication—people acquire precisely the same content that is in Bob's own locus. For example, you acquire the exact same top-spin service with a slight jump that Bob favors, and every slightest motion you make is an exact replica of Bob's serve. At the other extreme—causation of random changes—people rewrite the information in their locus such that it typically bears no resemblance to Bob's serve. Here, for example, you might 'write' into your tennis serve locus the idea that you should wave at mom when up to serve.

Please take note that I am not following the information in the brain here, although of course it is necessary for the process. What I am keeping track of is the actual behaviors, and I am completely ignoring the question of what particular information content in the brain may be causing them. The latter will not always be unimportant, but it does not concern me in the present analysis, and it is irrelevant to the points I will make. When I talk about 'replication failure,' what I mean here is the inability of the copier to *perform* a serve that is identical to Bob's.

Let us look first at the causation of random changes. This will look silly, but we cannot gain the proper insights until we examine the full spectrum of possibilities. As silly as it sounds, suppose I put 'wave at mom' in my tennis serve locus after watching Bob's top-spin serve. *You* will put randomly different, but typically equally dissimilar, information to Bob's serve in your own tennis serve locus (e.g. 'scratch the left knee'). What will happen? We are assuming that it is the content (i.e. the sequence of motions) involved in Bob's serve that make it attractive, in turn precipitating changes in the tennis-serve loci of other people. Given this, I myself (who now wave at mom when I 'serve')—and all others who randomly changed the information at their tennis serve loci after watching Bob—are not similarly beacons of change; our new 'tennis serves' look nothing like Bob's and they therefore get nobody excited (and mostly irritate the judge because they are not admissible). Bob's serve has not become more common, nor has the mean serve of the population moved in the direction of Bob's serve. Since evolution is about statistical changes in a population, the fact that this process does not produce reliable directional movement in the population's mean serve implies that this process cannot lead to cumulative design changes. After all, the first requirement for cumulative adaptive design is the possibility of directional change.

Now consider the other extreme. This will look silly too. Here, watching Bob's serve produces verbatim replicas in observers' tennis-serve loci. People copy perfectly, so there is never any mutation—not ever. What happens? Because Bob has the most attractive serve, all of the people who now have Bob's serve in turn become models for other people, who again copy the serve precisely and so forth. Bob's serve spreads until everybody is serving identically. Here, too, selection cannot lead to cumulative design changes because the serves have all become identical to Bob's. Nobody ever makes mistakes, so the future will be a world where every single one of us serves exactly like Bob—forever.

We thus see that at either end—totally random changes (zero copying fidelity), or perfect replication (100% copying fidelity)—there can be no accumulation of adaptive design. This can occur only somewhere 'in the middle', where descendant changes are relatively similar to the 'parent' stimulus, but somewhat different. There are two ways in which this can happen: (1) descendant serves are identical to the parent, but every once in a long while there will be a small accidental difference; or (2) the descendant serves are always accidentally different from the parent serve, but jump around relatively closely to the average of copying accuracy. In both cases we get more attractive future serves *by making marginal changes to Bob's, which in turn makes the marginally improved serve*

the new model (and this is what allows for cumulative adaptation). I examine each in turn.

(1) *Small copying mistakes only once in a long while.* Here the information 'written' in a person's tennis-serve locus is a replica of the 'parent' serve. There is a very small probability of replication failure so, *very rarely*, a random modification results. Such modifications will typically make Bob's serve less effective because a tennis serve is a complex behavior where many variables must be kept within narrow ranges to ensure success. I am assuming that only effective serves are attractive, and so most random changes will result in less attractive serves. But very, very occasionally, a random copying mistake begets a more effective—and therefore more attractive—serve, which then displaces Bob's as people now begin making replicas (i.e. perfect copies) of the improved serve. Many iterations of this cycle will lead to ever better serves.

I have just described a process of accumulation of adaptive design emerging from cultural transmission that is exactly parallel to cumulative genetic evolution by natural selection. Dan Sperber, as we saw, claims that in order for selection to produce cumulative design in cultural transmission, the process should look like this. But let us take a look at a rather different process.

(2) *Copying <u>always</u> involves mistakes, but around an average of perfect accuracy.* This process is illustrated below in **fig. 1.** Every time somebody sees Bob's top-spin serve, the *goal* is to copy it exactly, but there is always some error, and thus there is almost never a perfect copy. However, the errors are relatively small and not biased in any particular direction, so that Bob's serve is obviously the template for all descendant serves. In this scenario, replication is the occasional exception. However, the population's *mean serve* is still Bob's, even if no individual serve is a true replica. The errors amount to a constant introduction of modest variations, from which a serve superior to Bob's will emerge, and which then will become the new model serve—the new template to copy—for all of us. When that happens, this new serve becomes the new mean of the population, with a new cloud of error around it.

If we concentrate on the population mean, it is clear that cumulative design is taking place. This is not like genetic evolution by natural selection (where replication is very high fidelity), but it is certainly the accumulation of adaptive design due to selection (and it is faster than natural selection because variants are introduced in every copying attempt).

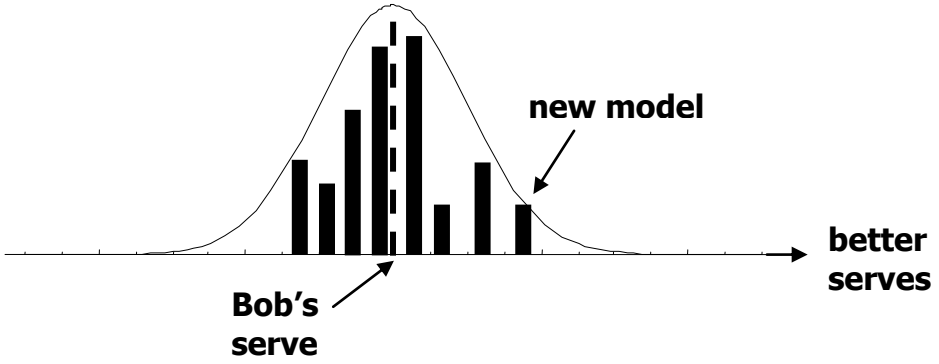

Fig. 1. Copying with modest errors. Think of the units in the X-axis as being very small, so that the distance between the left-most bar and the right-most bar is not too great—that is, we are assuming that all serves produced are minor deviations from the target serve (which is Bob's).

In the second case just considered replication rarely if ever happens; the norm is replication failure. It is a good summary description of the assumptions that go into many of the selectionist models that Boyd & Richerson (1985) introduced in their approach. This condition of replication failure as the norm is what Sperber claims renders cumulative adaptations from cultural transmission impossible. But we have just seen that it is certainly conceivable, and this lays bare that replication itself is a red herring. It is neither here nor there. What cumulative adaptation requires is (1) sufficient inaccuracy in the production of descendants such that superior variants can occasionally emerge; and (2) sufficient accuracy that, at the populational level (the mean), we can speak of meaningful, directional change (cf. Boyd & Richerson 2000).

E. Mutations may have consistent biases

But what about directed mutation? This idea posits an attractor, created—for example—by a psychological bias, towards which serves will tend because the copying mistakes we make are on average in the direction of the attractor. That is, the mean of our copying errors will not be zero. Contra Sperber, this is still not a problem—at least not *in principle*.

The attractor could be anywhere at all, but we can get our bearings by again considering the two extremes, namely, (1) when the attractor is the optimally effective serve, and (2) when it is in a direction *opposite* to the optimally effective serve.

> (1) *The mutation attractor is the optimally effective serve.* This case is illustrated below in **figure 2.** As before, suppose that every person tries to copy Bob's serve exactly, but fails within a cloud of error with mean zero. A few people, however, can see forward to the kinds of modifications that would make Bob's serve even

better, and attempt these. This means that the actual mean 'error' for the whole population will be skewed by these innovators in the direction of the optimal serve. Does this prevent cumulative adaptive design? No. On the contrary, it speeds up the process that takes the population to the optimal serve because mutations in this direction are slightly more likely. The design is cumulative because foresight does not extend to the optimal serve itself, merely to slight modifications of observable serves that take them in that direction.

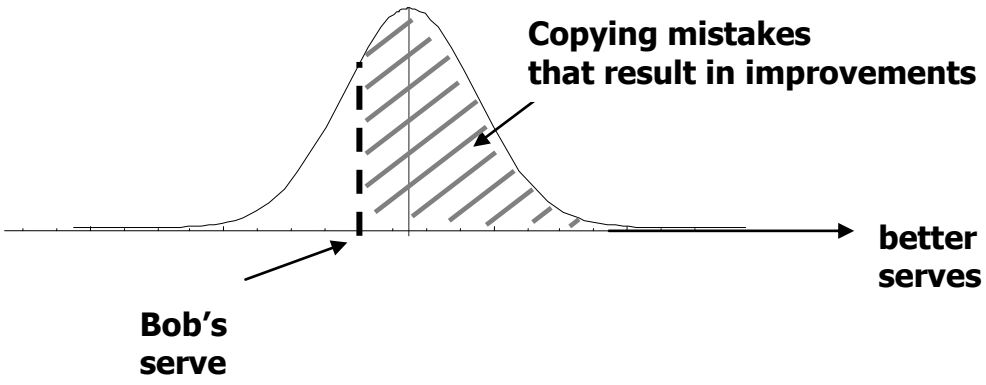

Fig. 2. Adaptive mutation bias. In this case the population mean is closer to the optimum, after copying, than is Bob's.

(2) *The mutation attractor is in a direction opposite to the optimal serve.* This case is illustrated below in **figure 3.** This could mean, for example, that there is something about the way it feels natural to move our bodies that makes us more likely to make errors in a direction away from the optimally most effective serve. But the phrase here is *more likely.* It doesn't mean that copying errors in the direction of a better serve never happen. Thus, what happens is that the mean copying effort results in a serve somewhat lower in quality than Bob's, but if the cloud of copying error occasionally produces a serve better than his, this serve will become the new target for copiers. This results in a new population mean that is again less good than the new target serve, *but it is not less good than the previous mean serve in the population.* Thus, the population mean will have moved closer to the optimal serve despite the fact that the mutation bias always makes it lag behind its current target.

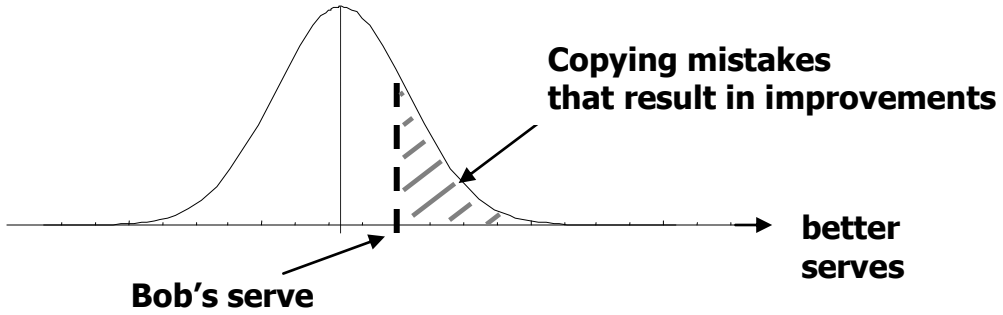

Fig. 3. Maladaptive mutation bias. In this case the population mean is further away from the optimum, after copying, than is Bob's serve. However, some copiers will make mistakes to the right of Bob, and since this yields a better serve, it will become the model for the next generation.

Only when the attractor is so far away that it prevents the emergence of *any* variants better than Bob's serve would the emergence of cumulative design be short-circuited, as shown below in **figure 4**.

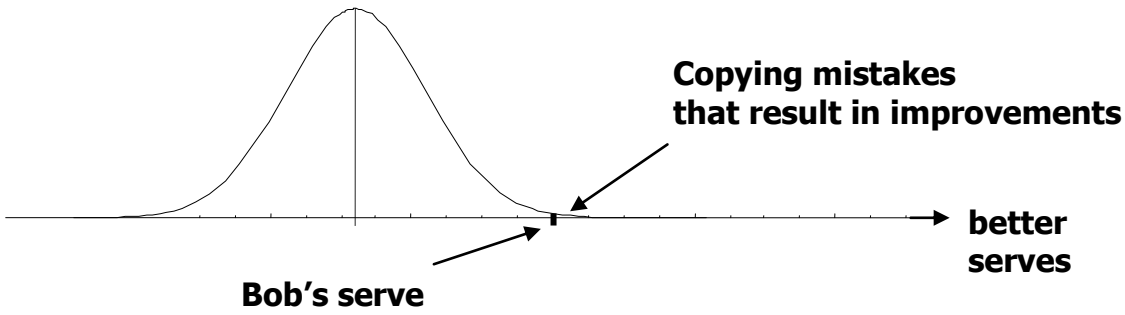

Fig. 4. Overly strong maladaptive bias. Due to a strong mutation attractor, the population mean is so far away from Bob's serve in a maladaptive direction that better serves practically will never appear.

The last example above shows that, when directed mutation occurs, it should be modeled together with selection. The direction of the system will then result from the algebraic sum of all the forces considered. We don't have to decide whether *either* mutation *or* selection is *the* force to consider in our modeling exercises. For problems having the structure just considered, Sperber will be right that constant, directed mutation, prevents cumulative adaptation *only if and when* such mutation is (1) not towards the optimum and, (2) of sufficient strength. This is an empirical question, and it may be true for some domains and not for others. But we will not find the answer under the armchair.

What about the evidence? Do we have *empirical* examples of cumulative cultural adaptations through selection? Yes. Other than tennis serves, we could name tennis racquets. In fact, we could name anything in the large domain called 'technology'. Here

design has obviously accumulated gradually. And even here Sperber's dictum that replication is a limiting case rather than the norm is correct (except in the case of our very modern manufacturing techniques).

One can also point to institutions. Certainly institutions have been 'constituting' themselves on paper for a long time, but institutional organization pre-dates paper. Moreover, though the rules of an institution may be written, institutional behavior is always in the (sometimes very) flexible neighborhood of what is written down, rather than a rigid instantiation of it. In this sense—as living, breathing organisms—institutions are always imperfectly copied (for an example, consider that the Mexican political constitution is—on paper—almost a replica of the US Constitution, on which it was modeled). And yet institutions accrete cumulative changes. The evidence that they do so adaptively is in the incontrovertible fact that complex societies have outcompeted simple ones, and in the fact that different institutional arrangements have been the key to success in the competition between different complex societies (McNeil 1963, Landes 1998, Diamond 1997, Wright 2000). Technological and institutional change are not the only examples, merely the most obvious ones. But they occupy much of what is important in cultural evolution, so they make the case that selectionist approaches will be quite significant to explaining culture.

Given that cumulative cultural adaptations don't require memes to replicate, this was not the litmus test for Darwinian analyses to culture. And if my critique of gene-analogy fetishism among the critics of 'memetics' is acceptable (for a mathematical demonstration of my core arguments, see Henrich and Boyd 2002), it simultaneously refutes the arguments of *proponents* such as Dawkins, Dennett, and Blackmore, who fetishize the alleged importance of 'replication' for opposite reasons.

'Imitation' is another red herring

A related point can be made about 'imitation' (i.e. what we do when we copy Bob's serve). Blackmore insists on imitation as *the* memetic process. But she would like to consider a narrative, for example, a 'meme.' And yet, narratives are not transmitted by imitation. Blackmore (1999:6) gets around this by corrupting the meaning of 'imitation' just as she did with 'replication':

> Dawkins said that memes jump from 'brain to brain via a process which, in the broad sense, can be called imitation' (1976:192). I will also use the term 'imitation' in the broad sense. So if, for example, a friend tells you a story and you remember the gist and pass it on to someone else then that counts as imitation.

With such a loose definition of 'imitation,' a reader such as myself cannot understand what standard Blackmore upholds when she insists that 'imitation' is what identifies the subject matter of 'memetics' (cf. Plotkin 2000:76-77).

But this is another red herring anyway. We need a handle on the social-learning cognitive mechanisms which, in combination with individual-learning processes, are responsible for affecting the distribution of memes (cf. Plotkin 2000; Laland & Odling Smee 2000). Imitation is important, but there is hardly any need to become obsessed with it. Different domains will involve different processes and will need mid-level theories particular to them, but "In every case the Darwinian population approach will illuminate the process…" (Boyd & Richerson 2000:144).

> The imitation of a motor act, the acquisition of a native language, and learning one's culture-specific social constructions have different developmental trajectories. . .Each is based on different psychological mechanisms. It is almost certainly the case that the characteristics each displays in terms of fecundity, longevity, and fidelity of copying are also different in each case, and different precisely because each is based on different mechanisms. The suggestion that "we stick to defining the [sic] meme as that which is passed on by imitation" (Blackmore 1998), if taken literally, is an impoverishment of memetics for reasons of wanting to maintain copying fidelity.—Plotkin (2000:76)

The insistence on imitation, as Plotkin suggests, comes precisely from this obsession with replication (copying fidelity). Imitation, narrowly (i.e. properly) understood, is the mechanism that strikes some observers, Blackmore included, as closest to the production of carbon copies. So they insist on the *word* 'imitation' because it confers the cachet of 'replication,' which in turn supposedly grants in exclusivity the legitimacy to undertake Darwinian analyses. Absurd.

And here again, the critics of 'memes' agree with this fetishism of 'imitation' only so they can reach the opposite conclusion. Atran (2001) in a section title, says, "No Replication without Imitation; Therefore, No Replication" (because there is no real imitation), and thus—absent replication—no applicability of Darwinian selectionist analyses to culture. This is hardly better, and refuting Blackmore's error is simultaneously to refute this one. If imitation and replication are neither here nor there when it comes to establishing a litmus test for the possibility of a Darwinism of culture, then one cannot reduce one's advocacy or skepticism of this project to whether there is or isn't imitation and/or replication.

It is true that some cultural transmission scholars have made much of 'imitation' (e.g. Boyd & Richerson 1985, 1996, 2000; Tomassello *et al.* 1993), and they have stressed its indispensability to cumulative cultural evolution. But these authors are interested in imitation as the ability which *initially* set humans along the path of cumulative cultural change (it is not a coincidence that when the above authors stress imitation they are comparing humans to nonhumans). But a phylogenetic question such as this one hardly denies that since the acquisition of imitative capacities other tricks have become possible for humans. For example, I have recently argued that language became possible when imitation led to the emergence of prestige hierarchies (Gil-White 2002). But this emergence of language now makes prestige-biased transmission often a process of

influence that pushes attitudes back and forth along a continuum, rather than imitation (Henrich & Gil-White 2001). Another example: narratives can accrue cumulative changes through selection, and I doubt that Robert Boyd, Peter Richerson, or Michael Tomassello will disagree. But narratives don't spread through imitation, even if the evolution of imitation was necessary for the emergence of language, which is indispensable for narrative. We must distinguish the phylogenetic indispensability of imitation from its current importance in cultural transmission.

Platonic inferences

So far I have ignored the following problem: although individuals do not make replicas of the memes they try to copy, they do *try to*. However, what could their target be? After all, our tennis player, Bob, never replicates his own serve perfectly either! Bob's performance is *itself* a cloud of error around a mean. So copiers must be abstracting an 'ideal Bob serve'—which they try to emulate—from Bob's performances. Sperber (1996:62-63) dismisses this as 'a Platonist approach' (indeed, as Dawkins [1999:xii] points out, Plato would have liked the argument that we strive to copy not the thing we see, but its 'essence', as we infer it, so to speak). To Sperber, formal properties cannot be causal.

I believe the opposite. It makes perfect sense that we infer and abstract an 'ideal' serve as Bob's goal, and then strive for it. For evolution to have designed our social-learning psychology otherwise would not have been adaptive, given that the performances of the people we copy are statistical clouds (cf. Dennett 1995:358; Dawkins 1999:x-xii; Blackmore 1999:51-52; Boyd & Richerson 2000). In a selectionist model it is therefore perfectly valid to define 'the meme' as the abstraction for which Bob strives, and to track the population mean as people try to copy this abstraction. I do not agree with criticisms that selectionist models have illegitimately relied on assumptions of discrete memes (Atran 2001), or that "the notion of replication certainly is one idealization too many for models of cultural transmission" (Boyer 1998). The problem being modeled will determine whether the simplification is legitimate, and many such models actually include copying error as a parameter anyway.

However, there is no question that there is an important role here for cognitive psychology and anthropology. We need a better understanding of how the brain decides which aspects of a performance are important and which irrelevant. Understanding such cognitive filters will tell us, for a particular domain, what is the 'meme.' But not having yet a good handle on such things is no obstacle (*pace* Atran 2002:ch.10) to current selectionist models (review in Feldman & Laland 1996) for these are concerned with the formal, emergent properties of Darwinian systems that, by assumption, are capable of cumulative adaptation, rather than with the histories of any specific, individual memes. As such, they teach us how to think about cultural evolutionary processes involving broadly specified—that is, relatively abstracted and idealized—memes, and the long run properties of dynamic systems having two interlocking systems of inheritance: genetic and cultural. What I have tried to do here is show that the assumption of selectionist

models—that cumulative adaptation is rampant in cultural transmission—is a very reasonable assumption.

What are the boundaries of 'a meme'?

Some critics (e.g. Atran 2001) accuse that memes don't have well-defined boundaries, but even "well-disposed" anthropologists can't see where to draw them. Maurice Bloch (2000) expresses his misgivings as follows:

> As I look at the work of meme enthusiasts, I find a ragbag of proposals for candidate memes, or what one would otherwise call units of human knowledge. At first, some seem convincing as discrete units: catchy tunes, folk tales, the taboo on shaving among Sikhs, Pythagoras's theorem, etc. However, on closer observation, even these more obvious 'units' lose their boundaries. Is it the whole tune or only a part of it which is the meme? The Sikh taboo is meaningless unless it is seen as part of Sikh religion and identity. Pythagoras' theorem is a part of geometry and could be divided into smaller units such as the concept of a triangle, angle, equivalence, etc.

Bloch has rather quickly pronounced defeat. These problems are hardly insurmountable, and they are not any different from similar conceptual problems faced in evolutionary genetics.

What is 'the meme'? The whole tune or only part of it? A Darwinian unit is of whatever size selection favors. This is why in evolutionary genetics Dawkins (1983:87-89) doesn't like to insist on the gene as a *cistron* ('start' codon to 'stop' codon). He is right. The cistron is more useful to molecular biologists. A tune, just like a cistron, has a starting point and an ending point, and, just like a cistron, this is a matter of *performance*, not selection. For the tune, a musical performance; for the cistron, the construction of a polypeptide chain. Our intuition that the *whole* tune is a unit does not come from an analysis of what people can remember and what they rebroadcast, about what spreads and doesn't spread *as a memory*, but rather from our understanding of the conventions of musical performances. The key point is that there are memes about *which* things to perform, and *how much* of them to perform, and these are of a different kind, and are found at different cultural 'loci,' than the loci which store tune fragments.

At one cultural 'locus' we find competing beliefs about which piece should be played. This locus can house a finite number of such beliefs, but the meme that says 'Beethoven's 5th deserves to be played' has consistently triumphed in securing a spot. A different locus is where memes compete to specify how much of a piece should be played. Here the belief 'play a piece from beginning to end' has fared well against competitors. Thus, it is because these two memes are successful in their respective loci that Beethoven's 5th is played often and in its entirety—not because the symphony itself is encoded whole in the heads of listeners! What listeners remember of the piece is stored in yet another locus where tune-fragments compete to be remembered. For the most part, only the opening theme to the 5th is encoded in this locus (because it is very catchy).

That these loci are independent (though not unrelated, of course) is made evident by the fact that very catchy but tiresome pop-tune fragments will get remembered so easily that the preference for the entire song *not* to be played will spread (at least after the initial success of the song in question). It is thus possible for the tune-fragment, on the one hand, and the negative preference for the song which contains it, on the other, to be simultaneously at high frequency, and remain so for a while. It was for a long time impossible for many of us who *hated* 'The Macarena'—a pop-tune that spread all around the world because of its linkage with a catchy dance routine—to get it out of our heads, but not because we liked it. We did not want to hear it played. Of course, for any tune-fragment to persist across the generations, a reasonable fraction of people must preserve the belief that the piece which contains it ought to be played. For this reason, the opening theme to Beethoven's 5th will probably continue to make it, but my future children will certainly never know 'The Macarena.'

What we have discovered here is that for a meme to spread—here, the opening theme to Beethoven's 5th—it needs a favorable ecology of other memes at other loci (for example, 'Beethoven's 5th deserves to be played'; the memes necessary to play a violin; the meme that violinists should be paid; etc., etc.). This discovery looks a lot like an earlier discovery: that any gene cannot hope to prosper unless it is surrounded by a favorable ecology of genes at other loci in its own organism, and also in the ecology of phenotypic effects of other organisms' genes. What else is new? If this discovery does not hurt the possibility of population analyses in biology, why should it be fatal for culture?

Yes, the Sikh taboo is more likely to spread and remain stable in an ecology of religious memes that are congruent with it. Yes, Pythagoras' theorem cannot be learned without first possessing the meme that says what a triangle is. But neither can the gene for reciprocity spread, for example, unless there are genes already for, say, social aggregation. None of this is new, or especially difficult.

Another vexing problem raised by the question "what are the boundaries of 'a meme'?" refers to the level of abstraction. When somebody tells me a story, and I retell it, I will never give a verbatim rendition of the story I heard. Many of the details will change. There are good reasons to think that most of the details are not even stored in memory (Schank & Abelson 1995). I can feel the critic pouncing: "Aha! There is no stability!" But at what level? Suppose that the *skeleton* of the story is very stable. If so, the fact that story details are not even encoded in the listener's brain—and therefore change radically from version to version—is as worrisome to Darwinian analyses in culture as silent mutations in the DNA code—when the mutant base codes for the same thing as the previous base—are to evolutionary genetics (i.e. not at all). What we need to keep track of is the story skeleton. Changes *there* will be the real mutations.

Meme 'content' is not everything

Sperber (2000) recently made a concession to the point that we make Platonic inferences but then insisted that these are almost always *triggered* rather than *bootstrapped.* Atran (2002, 2001, 1998), and Boyer (1998, 1994) make essentially the same point.

The argument is that observation produces 'inferences' which are best described as the triggering of a pre-existing knowledge structure. Sperber (2000:165-66) gives the example of language, interpreted from a Chomskian point of view, "where language learners converge on similar meanings on the basis of weak evidence provided by words used in an endless diversity of contexts and with various degrees of literalness or figurativeness." From this it follows, he says, that language learning is much more about triggering *pre-existing* knowledge rather than bootstrapping *new* knowledge. There are no stable and discrete memes competing with each other in a selective contest, goes the argument, but memes mutating quickly and fuzzily, morphing inexorably into the shape favored by the content-bias 'attractors' which our innate cognitive endowment specifies.

Sperber admits that not everything is like that. "Learning to tap dance involves more copying than learning to walk," but, he insists, "For memetics to be a reasonable research programme, it should be the case that copying [as opposed to triggering], 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." That is not the case, he says. Instead, "the acquisition of cultural knowledge and know-how is made possible and partly shaped by evolved domain-specific competencies…"

In my view Sperber is setting up a straw man—a false test—for several reasons. First, because, he is asking us to choose between complements rather than between alternatives. If we have domain-specific competencies this hardly rules out the possibility of rampant selection-driven cumulative adaptations.

Second, because, for a great many domains the triggering of inferences makes a rather different point. Our toy example will assist us here too. As observed earlier, learning Bob's serve requires that we abstract his *goal* from his statistical cloud of performances. This is an *inference*, sure, and it relies on "pre-existing knowledge" too. But knowledge about what? Primarily, about the purpose of a serve in a game of tennis. In other words, knowledge that does not come from an innate, domain-specific module as Sperber would have it, for the brain of a human hardly comes prepared to trigger "tennis" (and many people around the world don't play it). An important form of cumulative bootstrapping takes place already merely in the fact that the rules of tennis need to be understood *first* in order properly to infer the specific thing that Bob is going for when he serves. There is no straightforward or absolute reduction to the triggering of innate modules here.

Third, because Sperber's linguistic example is not even very good. Although there is undoubtedly much innate knowledge dedicated to the bootstrapping of language, a model that reduces linguistic historical processes to nothing more than the triggering of innate knowledge can never explain how the common ancestor of Indian and European

languages (whether this was Sanskrit or the alleged 'Indo-European') became Hindi in one place and Spanish in another.

Fourth, because Sperber's test is unfairly asymmetric. In his formulation (see the quote above), the mechanism he does not favor—the copying of knowledge—must be "overwhelmingly" dominant, and yet his own favored explanation need only be "partly" responsible for his prescriptions to triumph. Tails, he wins; heads, we lose!

Finally, even should we grant all of Sperber's assumptions and accept that all attractors will be innate, and that there will be attractors for everything, he is still wrong. Henrich & Boyd (2002) show that so long as more than one attractor can exert influence over a given meme, and the attractors are strong relative to selection pressures, the dynamics quickly become a contest between the discrete alternatives favored by each attractor, engaged in a selective contest. So even here the fuzzily-morphing-into-the-attractor model is not right—something close to particulate selection still happens.

F. Non-content biases and their importance

The last line of defense for Sperber would then be that, even so, the contest is all between innate attractors and so one cannot expect cumulative cultural evolution acting on *arbitrarily* varying memes. Atran (1998) and Boyer (1998) agree with this view that transmission is mostly about moderate variations around 'core memes,' which are strongly constrained by innate mental biases that focus on a meme's *content*. A related view has stressed that the main causes of 'triggered inferences' will be local non-cultural environments (e.g. Tooby & Cosmides 1992), so cultural differences reduce to the environmental conditions surrounding the various local human populations. However, others argue—not *in stead* (content biases are important too) but *in addition*—for the importance of *non*-content biases that allow arbitrary differences to spread and remain stable (Boyd & Richerson 1985; Henrich & Boyd 1998; Henrich & Gil-White 2001; Gil-White 2001a, 2001b). To see why we believe in the rampant spread of arbitrary differences, we must describe the relevant social-learning cognitive biases.

Assume that Bob is your hero because he is a great tennis player. Bob likes a Wilson racquet. What do you do? Buy a Wilson racquet. Bob wears leather pants; you buy leather pants. Or suppose everybody in your high school class is getting leather pants. What do you do? Get leather pants (you don't want to look like a deviant). In these examples you acquire the meme not because the meme itself captivates you; what seduces you are the contingently associated features: the meme's source, or its relative frequency. In these observations lies a key—and very misunderstood—virtue of the selectionist approach pursued in the tradition pioneered by Boyd & Richerson (1985): the importance of 'non-content' transmission biases.

The memes that do well and spread widely in a population are those which, for whatever reason, the human brain has a 'taste' for. But, as seen above, some of these 'tastes' may have nothing to do with the actual content of a meme (what the meme actually 'says', 'prescribes', or makes people do). Of course, many biases involved in social learning will

focus on a meme's content (I am calling them 'content biases,' but Boyd and Richerson call them 'direct biases'). However, as students of culture from an anthropological perspective, Boyd and Richerson have devoted much attention to the long-term consequences of non-content biases that can cause the accumulation of arbitrary differences between societies. The non-content biases relevant to this problem are conformity bias and prestige bias.

Much research in social psychology suggests that humans have biases to prefer memes that are common relative to competing memes at a particular cultural 'locus' (Miller and McFarland 1991; Kuran 1995; Asch 1956, 1963[1951]). Boyd and Richerson (1985:ch.7) and Henrich & Boyd (1998) give models to explain the adaptiveness of informational conformism as helping individuals pick up useful memes that others have already converged on. Gil-White (2001a) argues that interactional-norm conformism is adaptive because it gains the conformist the maximal number of potential interactants.

Boyd and Richerson have also speculated (as indeed have many others) that prestigious individuals are copied more often than others, and Henrich & Gil-White (2001) recently took these speculations and developed a lay model to explain the evolution of such a cognitive bias, reviewing also the evidence for it extant in the social-scientific literature. We argue that prestige-bias is adaptive because successful individuals (i.e. with better memes) tend to have prestige.

These two biases care nothing about content: conformity bias cares about relative frequency, and prestige bias about source. As far as these biases are concerned, the memes could be 'about' anything at all. Thus, in domains without strong content biases, we should see the following effects. First, the memes of prestigious individuals will tend to become more common, but these will be unpredictably different for people in different communities given that every individual has an idiosyncratic life history (e.g., I, but not you, may fall off the horse after washing my feet in a stream, and conclude superstitiously that the stream was somehow directly responsible). Second, such differences will be larger between members of different communities (even if we both fall off our horses after washing in the stream, I am more likely to blame the stream if my local community already believes that streams have supernatural powers). This sort of process will engender arbitrary differences between societies, and a third effect—conformism—will keep them stable at high frequency (in a community as large as the sampling area for the conformist bias). The fourth and last effect is historical: such stable differences between societies in turn become acquired 'content biases' acting on the evolution of future memes, setting different societies on separate and distinct historical paths.

The conformist and prestige biases therefore offer themselves as an appealing joint explanation for the different historical trajectories which have caused dramatic variation among the world's cultures. (Drift can also act to bootstrap arbitrary differences to frequencies high enough for conformity to kick in and stabilize them.) Together they can explain why two populations living in the same environment could become quite different, culturally—something that happens all the time.

G. Don't reduce everything to 'content'

The issue of cultural variability has been an anthropological concern throughout the 20th century, and it has led to the theoretical excess of 'cultural relativism,' which holds that human brains are—for any and all purposes—blank-slates upon which a local culture can write literally anything at all. That this is false should have been obvious (regrettably, it hasn't been). But perhaps some anthropologists are now guilty of overreacting in claiming that the blank-slate view of culture is *always* wrong.

> The picture of the human mind/brain as a blank slate on which different cultures freely inscribe their own world-view. . .[is] incompatible with our current understanding of biology and psychology.
>
> . . . the brain contains many sub-mechanisms, or 'modules', which evolved as adaptations to. . .[ancestral] environmental opportunities and challenges (Cosmides & Tooby 1987, 1994; Tooby and Cosmides 1989, 1992) [and]…are crucial factors in cultural attraction. They tend to fix a lot of cultural content in and around the cognitive domain the processing of which they specialize in.—Sperber (1996:113)

Other anthropologists in this tradition have expressed similar views in the process of exploring some interesting content biases as the reason for the widespread recurrence of certain memes. For Boyer (1994) these are certain religious ideas; for Atran (1998) concepts of living-kinds; and for Hirschfeld (1996), intuitions about so-called 'races.' These are all valuable enterprises, but these authors seem to think that the discovery of these content biases amounts to a refutation of the possibility of acquiring any unconstrained memes (Boyer 1998), and therefore a refutation of the possibility of stable, arbitrary differences between cultures (Hirschfeld 1996:21-22), which in turn implies a refutation that such nonexistent differences could lead to cultural group selection (Atran 2002:ch.10). One should not conclude that finding content biases in some domains excludes the possibility of arbitrary differences in other domains without strong content biases. Sperber seems to present the issue above as an either/or question: the brain is not a blank slate, therefore cultural content is fixed around the cognitive domain of our evolved biases. But we must adjudicate this on a domain-by-domain basis. The blank-slate assumption may in fact be a reasonable approximation in a great many domains.

With a different slant, Blackmore (1999) and Dennett (1995) also argue for the primacy of *content,* but they place the focus on the meme, rather than on innate psychology. Cultural evolution is here a selective process that makes memes increasingly better propagators. As Dennett (1995:362) puts it,

> Dawkins (1976:214) points out that '…a cultural trait may have evolved in the way it has simply because it is advantageous to itself.' (…)
> The first rule of memes, as for genes, is that replication is not necessarily for the good of anything; replicators flourish that are good at…replicating—for whatever reason!

In other words, memes that 'look' like what the brain 'wants' will spread even if they lack the effects that the brain is adaptively 'hoping for.' This is valid, but the emphasis on *content* as such is overplayed.

Dennett and Dawkins suggest that the only thing affecting a meme's spread is whether *the meme itself* is good at replicating, and that selection will successively edit the meme's content so that it is ever better at replicating. This is the 'meme's eye view': only the properties of a meme (i.e. its content) determine its spread. But a meme can be lucky. It can happen to find itself in the head of a prestigious person, or, thanks to prestige-bias bootstrapping (or even random drift processes), it may find itself at high frequency through no 'fault' of its own. In both cases the meme's content takes a back seat. In fact, the meme may be favored *despite* its content. This means that prestige-biased and conformist transmission are excellent explanations for why some maladaptive memes spread and remain stable, even when the memes themselves are *not* good at replicating. I hardly think that Dennett's 'first rule of memes' is a rule at all, let alone the first. It is in no way necessary as an all-encompassing perspective on the processes involved in cultural transmission.

I am hardly alone in making this criticism (e.g. Conte 2000:88; Laland & Odling Smee 2000:134; Boyd & Richerson 2000), and I am hopeful that the authors criticized here can be convinced. After all, Atran (2002:ch.10), partially acknowledges that "from a cognitive standpoint, some cultural aspects are almost wholly arbitrary." Boyer (1998) recognizes the importance of prestige bias, and Sperber (1996:90-91) explicitly recognizes its power to generate arbitrary differences between societies. Meanwhile, Blackmore (1999:ch.6) talks about source biases that I don't believe exist (e.g. 'imitate the good imitators') but which, as source biases, should undermine her view of meme-selection as solely the result of meme content. Dawkins (1999:vii) starts his introduction to Blackmore's book by describing prestige bias. And Dennett and Dawkins are clearly aware of frequency-dependent effects such as conformism (Dennett 1995:352). Following these authors' own observations about non-content biases to their logical conclusions entails that arbitrary differences between cultures are not only possible but likely, and to the extent that they are stable they generate selection pressures at the group level (Boyd & Richerson 1985:ch.7; Henrich & Boyd 1998).

We can now closely evaluate the sometimes facile claims made about memes, whether by proponents or critics. Susan Blackmore (1999, 2000) has recently become the most outspoken proponent of the notions I have just criticized, although the main points are owed to Dennett (1995) and also to Dawkins (1989). Her most pithy formulation, and the one that makes all of her intended links, is the following (Blackmore 2000:26):

> …memes clearly vary and therefore fit neatly into the evolutionary algorithm. In other words, memes are replicators. The importance of this is that replicators are the ultimate beneficiaries of any evolutionary process. Dennett (1995) urges us always to ask *cui bono*? or who benefits? And the answer is the replicators…

I believe every link in this argument to be mistaken. Blackmore begins by saying that it is *because* memes vary that they fit into the evolutionary algorithm. But this is false. Grains of sand vary, and they do not fit into the evolutionary algorithm. Memes fit into the algorithm only *if* they vary *and* remain reasonably stable in the process of transmission. If mutation is so great in both frequency and magnitude that selection cannot act on the variation, then "what is passed on in imitation" (how Blackmore [2000:25] defines 'memes') would certainly *vary* but it could not be analyzable with Darwinian tools.

Second, Blackmore says that *because* memes fit into the evolutionary algorithm, they must be replicators. This again is false. Units can fit into the evolutionary algorithm even if they don't replicate, as I have argued with the example of Bob's tennis serve. Boyd & Richerson (1985:ch.3) already demonstrated long ago that this is true even for the case of blending inheritance (though hardly anyone ever seems to notice). Recently, Henrich & Boyd (2002) provide another demonstration of why replication is a red herring.

Third, Blackmore argues that *because* memes are supposedly replicators, and because "replicators are the ultimate beneficiaries of any evolutionary process," our analyses must always be in terms of how the memes benefit. False again, as shown by the existence of non-content biases. But this last argument of Blackmore's is so 'sexy'—it is responsible for most of the attention which her work, and the preceding work of Dennett and Dawkins has received—that it is worth a thorough refutation, which I turn to next.

Memetic Drive—the 'meme's eye view' gone mad

Can we reduce everything ultimately to the interests of 'memes'? Blackmore (1999:8) says that "We humans. . .have become just the physical 'hosts' needed for the memes to get around." But this would mean that, just as a chicken is an egg's way of making another egg (the 'selfish gene' perspective), a brain is just a meme's way of making another meme (the 'selfish meme'). Is this really what Blackmore is saying? Apparently. This is Blackmore's concept of 'memetic drive,' which, according to some (Aunger 2000:11) is her most radical idea:

> Genes are instructions for making proteins, stored in the cells of the body and passed on in reproduction. Their competition drives the evolution of the biological world. Memes are instructions for carrying out behavior, stored in brains (or other objects) and passed on by imitation. Their competition drives the evolution of the mind.—Blackmore (1999:17)

Since competition among genes is responsible for the emergence of biological design, and since the brain is a biological organ, one may hazard that by 'mind' Blackmore does not simply mean 'brain.' What she could cogently mean is that 'mind' is the set of interconnections that *end up* instantiated in the brain as the result of some developmental process which involves cultural inputs. In other words, the mind is partly a bunch of

memes—*partly*, because not everything that ends up instantiated in the brain is acquired socially, as some of it is innately given. So let us make the best case for Blackmore's argument, and artificially restrict 'mind' to "connections that result from the social acquisition of information." Can we now say that competition among memes drives the evolution of the mind in the same way that competition among genes drives the evolution of the biological world?

Yes. If we define 'mind' as whatever memes end up in a brain, then, tautologically, competition among memes drives the evolution of minds. The tautology is not entirely useless because the meme concept emphasizes Darwinian processes that have been neglected. But it is better to say it without a tautology which, to boot, requires a new technical definition of 'mind' (we must resist the temptation to 'conquer' sexy terms by redefinition). Better to say: "*short term* cultural evolution is the product of competition among memes because a 'culture' is a distribution of memes." Is this a truly new or radical argument? Certainly not by the standards of cultural transmission theory, relative to which Blackmore (1999:15-17) believes she has advanced so much that her ideas are in fact christening an entirely new and autonomous discipline which has yet to begin.

But perhaps my translation above is not adequate and Blackmore, by 'mind,' really *does* mean 'brain.' In chapter six of her 1999 book she actually argues that memes selected for big brains to serve their own—the memes'— 'interests' (what she calls 'memetic drive'). In other words, the 'interests' of memes set processes in motion that select for genes, which in turn code for brains that prefer those same memes. In this way, the memes are furthering their own propagative interests. A brain, we might say, is quite *literally* just a meme's way of making another meme.

This *is* radical. It's also wrong. A meme can select for a gene only if it is widespread (meta-populationally) and stable (inter-generationally), and there are but two avenues for such a condition to emerge. In the first, the meme is selected by an innate 'content bias' in the brain's design, making it widespread in the species and stable across time. But for Blackmore this is a catch-22, because what has put this meme in a position to select for a gene is the fact that this same gene evolved first. The meme is not in the "driver's seat."

The second avenue is if a process such as group selection through conformist transmission (Boyd & Richerson 1985; Henrich & Boyd 1998) makes a meme widespread and stable, even though there was originally no innate content bias to prefer it. For example imagine that group selection makes a meme for group-welfare altruism spread as groups which become more competitive (as a result of having this meme) displace others. If we suppose that another meme emerges in some of these groups that says 'punish non-altruists in your group,' these new groups will become even more competitive, and will displace groups with voluntary cooperation. Once such groups populate the world, it will become costly, *everywhere*, not to acquire the altruism meme quickly and reliably, because then one will incur punishment. In such a case, genes coding for an innate content bias specific to this meme—that is, an innate bias that makes it easy to acquire the group-welfare altruism meme—will be favored. Memes will have selected for genes in a Baldwinian process.

This can certainly work, but it is not radical by the standards of cultural transmission theorists, some of whom have been pushing this sort of argument for years (e.g. Boyd & Richerson 1985), much less does it call for an entirely new discipline. The reason is that it does not, as Blackmore claims, put the memes in the "driver's seat" to the detriment of the 'interests' of the genes when it comes to brain design.

Blackmore's arguments fail because she has not considered the complications involved in specifying the time-scales in which particular evolutionary processes take place. In the long run, as the economists say, we are all dead, and for this reason we must specify the time-scale. One must not confuse the true statement that the replicative 'interests' of memes affect *short term* cultural evolution, with the false statement that the replicative 'interests' of memes—*against* the 'interests' of genes—drive the longer-term process of brain design. The brain simply cannot be designed against the 'interests' of genes, because the design of the brain will be coded for by genes. And genes that alter the design of the brain cannot emerge without differential reproductive success in their favor. This is elementary. When memes select for genes it will be *only because* the 'interests' of memes and genes *coincide*. Granted, they may only coincide after the meme has become widespread (as in the example above) through short-term Baldwinian and culture-driven processes (and this is very interesting), *but they will still have to coincide*. And a coincidence is just that—it is not a radical "turning of the tables" on our understanding of what shapes brains, as Blackmore would have it.

The design of the brain will still be about biological reproductive success in the environments that selected for this design, not for the propagative success of memes in the utter absence of a biological instrumentality. Let us stop worrying about non questions based on false observations, such as "We seem to have a brain 'surplus to requirements, surplus to adaptive needs' (Cronin 1991:355)," and, ". . .our abilities are out of line with those of other living creatures and they do not seem obviously designed for survival" (Blackmore 1999:67-68).

Conclusion

I conclude by listing the morals. The first is that we need, not narrowly *genetic* Darwinian thinking, but a 'population thinking' attitude that considers—in its own terms—the properties of statistical populations capable of inheritance and subject to selection (Boyd & Richerson 2000). A narrow comparison of the details of genes and memes is not the right test, though there is hardly any reason to abandon the heuristic horsepower of the analogy.

The second moral is that if we believe psychological biases are the main source of selective forces acting on memes, then the discovery and implications of non-content biases should be taken seriously. This detracts nothing from the importance of content biases, it merely adds to the repertoire of forces that must be considered.

The third moral is that we have talked quite enough. One reason there is this much misunderstanding about what memes can or cannot be, and what they must or must not be for Darwinian analyses to apply, is that psychologists and anthropologists know so little evolutionary genetics, and this is not easy to remedy. But another reason is that psychologists and anthropologists have done very little to advance something they are eminently qualified to do: analyze the natural histories of particular memes in different domains, and the proximate cognitive biases responsible for such processes. Some of the points I have made here came to me as revelations after tracing the spread of one particular meme in the communities I study in western Mongolia (Gil-White, *in prep.*), and others as a result of trying to give a full account of one particular social-learning bias (Henrich & Gil-White 2001). More revelations will follow, as in any science. But, as in any science, we need to resist the pleasures of navel-gazing in the armchair in order to get our hands dirty and toil at the empirical problems.

References

Asch, S. E. (1956) Studies of independence and conformity: I. Minority of one against a unanimous majority. *Psychological Monographs* 70: (Whole No. 416)

Asch, S. E. (1963 (1951)) Effects of group pressure upon the modification and distortion of judgments. In: *Groups, leadership, and men*, ed. H. Guetzkow, New York: Russell & Russell.

Atran, S. (1998) Folk-biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and brain sciences* 21: 547-609

Atran, S. (2001) The trouble with memes: Inference versus imitation in cultural creation. *Human Nature* 12: 351-381

Atran, S. (2002) *In gods we trust: The evolutionary landscape of religion*, New York: Oxford University Press.

Atran, S., Medin D., Ross N., Lynch E., Vapnarsky V., Ucan Ek' E., Coley J., Timura C., Baran M. (2002) Folkecology, cultural epidemiology, and the spirit of the commons: A garden experiment in the Maya lowlands. *in press, Current Anthropology*

Aunger, R. (2000) Introduction. In: *Darwinizing culture: The status of memetics as a science*, ed. R. Aunger, Oxford & New York: Oxford University Press.

Blackmore, S. (1999) *The meme machine*, Oxford: Oxford University Press.

Blackmore, S. (2000) The meme's eye view. In: *Darwinizing culture: The status of memetics as a science*, ed. R. Aunger, Oxford & New York: Oxford University Press.

Bloch, M. (2000) A well-disposed anthropologist's problems with memes. In: *Darwinizing culture: The status of memetics as a science*, ed. R. Aunger, Oxford & New York: Oxford University Press.

Boyd, R., Richerson P. J. (1985) *Culture and the evolutionary process*, Chicago: University of Chicago Press.

Boyd, R., Richerson P. J. (1996) Why culture is common, but cultural evolution is rare. *Proceedings of the British academy* 88: 77-93

Boyd, R., Richerson P. J. (2000) Memes: Universal acid or a better mousetrap? In: *Darwinizing culture: The status of memetics as a science*, ed. R. Aunger, Oxford & New York: Oxford University Press.

Boyer, P. (1994) *The naturalness of religious ideas*, Berkeley: University of California Press.

Boyer, P. (1998) Cognitive tracks of cultural inheritance: How evolved intuitive ontology governs cultural transmission. *American Anthropologist* 100: 876-889

Castro L. and Toro, M. A. (2002). Cultural Transmission and the Capacity to Approve or Disapprove of Offspring's Behaviour. *Journal of Memetics - Evolutionary Models of Information Transmission*, 6. http://jom-emit.cfpm.org/2002/vol6/castro_l&toro_ma.html

Cavalli-Sforza, L. L., Feldman M. (1981) *Cultural transmission and evolution*, Princeton: Princeton University Press.

Conte, R. (2000) Memes through social minds. In: *Darwinizing culture: The status of memetics as a science*, ed. R. Aunger, Oxford & New York: Oxford University Press.

Cosmides, L., Tooby J. (1987) From evolution to behavior: Evolutionary psychology as the missing link. In: *The latest on the best: Essays on evolution and optimality*, ed. J. Dupré, Cambridge MA: MIT Press.

Cosmides, L., Tooby J. (1994) Origins of domain-specificity: The evolution of functional organization. In: *Mapping the mind: Domain-specificity in cognition and culture*, eds. L. A. Hirschfeld and S. A. Gelman, New York: Cambridge University Press.

Cronin, H. (1991) *The ant and the peacock*, Cambridge: Cambridge University Press.

Dawkins, R. (1982) *The extended phenotype*, Oxford: Oxford University Press.

Dawkins, R. (1999) Foreword. In: *The meme machine*, ed. S. Blackmore, Oxford: Oxford University Press.

Dawkins, R. (1989(1976)) *The selfish gene*, Oxford and New York: Oxford University Press.

Dennett, D. C. (1995) *Darwin's dangerous idea: Evolution and the meanings of life*, New York: Simon and Schuster.

Diamond, J. M. (1997) *Guns, germs, and steel : the fates of human societies*, New York: W.W. Norton.

Feldman, M. W., Laland K. N. (1996) Gene-culture coevolutionary theory. *Trends in evology and evolution* 11: 453-7

Gil-White, F. J. (2001a) Are ethnic groups biological 'species' to the human brain?: Essentialism in our cognition of some social categories. *Current Anthropology* 42: 515-554

Gil-White, F. J. (2001b) L'evolution culturelle a-t-elle des règles? *La rechérche* Hors-Série No. 5: 92-97

Gil-White, F. J. (2002a) Comment on Atran et al (2002). *Current Anthropology* 43: 441-442

Gil-White, F. J. (2002b) The evolution of prestige explains the evolution of reference, paper delivered at the *Fourth international conference on the evolution of language* (Harvard University)

Gil-White, F. J. *in prep.* I killed a one-eyed marmot: Why some narrative memes spread better than others, and how they maintain beliefs

Henrich, J., Boyd R. (1998) The evolution of conformist transmission and the emergence of between-group differences. *Evolution and Human Behavior* 19: 215-241

Henrich, J., Boyd R. (2002) On modeling cognition and culture: How formal models of social learning can inform our understanding of cultural evolution. *under review, Cognition and Culture*

Henrich, J., Gil-White F. J. (2001) The evolution of prestige: Freely conferred status as a mechanism for enhancing the benefits of cultural transmission. *Evolution and human behavior* 22: 165-196

Hirschfeld, L. (1988) On acquiring social categories: Cognitive development and anthropological wisdom. *Man* 23: 611-638

Hull, D. (2000) Taking memetics seriously: Memetics will be what we make it. In: *Darwinizing culture: The status of memetics as a science*, ed. R. Aunger, Oxford & New York: Oxford University Press.

Kuran, T. (1995) *Private truths, public lies: The social consequences of preference falsification*, Cambridge, MA: Harvard University Press.

Laland, K. N., Odling-Smee J. (2000) The evolution of the meme. In: *Darwinizing culture: The status of memetics as a science*, ed. R. Aunger, Oxford & New York: Oxford University Press.

Landes, D. S. (1998) *The wealth and poverty of nations: Why some are so rich and some so poor*, New York: W.W. Norton.

Lumsden, C., Wilson E. O. (1981) *Genes, mind, and culture: the coevolutionary process*, Cambridge, MA, and London: Harvard University Press.

McNeill, W. H. (1963) *The rise of the West: A history of the human community*, Chicago: University of Chicago Press.

Miller, D. T., Mcfarland C. (1991) Why social comparison goes awry: The case of pluralistic ignorance. In: *Social comparison: contemporary theory and research.*, eds. J. Suls and T. Ashby, Hillsdale, N.J.: L. Erlbaum Associates.

Plotkin, H. (2000) Memes through social minds. In: *Darwinizing culture: The status of memetics as a science*, ed. R. Aunger, Oxford & New York: Oxford University Press.

Schank, R. C., Abelson R. P. (1995) Knowledge and memory: The real story. In: *Knowledge and memory: The real story*, ed. R. S. Wyer, Hillsdale, NJ: Lawrence Erlbaum Associates.

Sperber, D. (1996) *Explaining culture: A naturalistic approach*, Oxford: Blackwell.

Sperber, D. (2000) An objection to the memetic approach to culture. In: *Darwinizing culture: The status of memetics as a science*, ed. R. Aunger, Oxford & New York: Oxford University Press.

Tomasello, M., Kruger A. C., Ratner H. H. (1993) Cultural learning. *Brain and Behavioral Science* 16: 95-552

Tooby, J., Cosmides L. (1989) Evolutionary psychology and the generation of culture, Part I: theoretical considerations. *Ethology and sociobiology* 10: 29-49

Tooby, J., Cosmides L. (1992) The psychological foundations of culture. In: *The adapted mind: Evolutionary psychology and the generation of culture.*, eds. J. H. Barkow, L. Cosmides and J. Tooby, New York and Oxford: Oxford University Press.

Wilkins, J. S. (1998) What's in a meme? Reflections from the perspective of the history and philosophy of evolutionary biology. *Journal of memetics--Evolutionary models of information transmission* 2

Williams, G. C. (1966) *Adaptation and natural selection*, Princeton, NJ: Princeton University Press.

Wright, R. (2000) *NonZero: The logic of human destiny*, New York: Pantheon.

[1] Unlike Sperber (2000:163) I don't think there is anything trivial about this definition, and neither do I think that it corresponds to how anthropologists have always thought about culture, as he claims. Implicit in this definition is the idea that memes are units, that they are materially stored, and that they are subject to selection. These intuitions open the way to a completely different form of analysis of culture from that which we anthropologists had been traditionally contemplating. As Sperber (1996) himself has repeatedly accused, anthropologists have been prone to mystical approaches to culture that put it 'out there' in the ether somewhere rather than in people's brains, and they have failed to examine the processes of transmission in its phenomenal and cognitive details. Making the units of cultural transmission analogous to genes, however loosely, which is what the 'meme' idea in any of its forms does, produces an entirely new perspective—in fact, a revolution of sorts.