An objection to the memetic approach to culture

Dan Sperber

Memetics is one possible evolutionary approach to the study of culture. Boyd and Richerson's models (1985, Boyd this volume), or my epidemiology of representations (1985, 1996), are among other possible evolutionary approaches inspired in various ways by Darwin. Memetics however, is, by its very simplicity, particularly attractive.

The memetic approach is based on the claim that culture is made of memes. If one takes the notion of a meme in the strong sense intended by Richard Dawkins (1976, 1982), this is indeed an interesting and challenging claim. On the other hand, if one were to define 'meme', as does the Oxford English Dictionary, as 'an element of culture that may be considered to be passed on by non-genetic means', then the claim that culture is made of memes would be a mere rewording of a most common idea: anthropologists have always considered culture as that which is transmitted in a human group by non-genetic means.

Richard Dawkins defines 'memes' as cultural replicators propagated through imitation, undergoing a process of selection, and standing to be selected not because they benefit their human carriers, but because they benefit themselves. Are non-biological replicators such as memes theoretically possible? Yes, surely. The very idea of non-biological replicators, and the argument that the Darwinian model of selection is not limited to the strictly biological are already, by themselves, of theoretical interest. This would be so even if, actually, there were no memes. Anyhow, there are clear cases of actual memes, though much fewer than is often thought. Chain-letters, for instance, fit the definition. The very content of these letters, with threats to those who ignore them and promises to those who copy and send them, contributes to their being copied and sent again and again. Chain-letters do not benefit the people who copy them, they benefit their own propagation. Moreover, some chain-letters are doing better than others because of the greater effectiveness of their content in causing replication.

Once the general idea of a meme is understood—and especially if it understood fairly loosely—it is all too easy to see human social life as teeming with memes. Aren't, for instance, religious ideas, with their threats of hell for unbelievers and promises of paradise for the proselytes, comparable to chain-letters, and in fact much more effective in benefiting their own propagation, come what may to their human carriers? More generally, aren't words, songs, fashions, political ideals, cooking recipes, ethnic prejudices, folktales, and just about everything cultural, items that get copied again and again, with the more successful items managing to invade more minds over longer periods of historical time, and to recruit those minds to further their own propagation? lf this were so, if culture were made of memes in Dawkins's strong sense, then the study of culture could-and arguably should-be recast as a science of memes or 'memetics'. The Darwinian model of selection could be used, with proper adjustments, to explain the properties, the variety and the solution of culture, just as it explains the properties, the variety; and the evolution of life.

The question is whether the claim that culture is made of memes is a true one. Several objections have been made to this claim. In his 'Foreword' to Susan Blackmore's *The Meme Machine* (1999), Richard Dawkins responds to the simplest and most serious objection, 'that memes, if they exist at all, are transmitted with too low fidelity to perform a gene-like role in any realistically Darwinian selection process' (Dawkins 1999: x).[1] I want here to discuss Dawkins's responses, and, in so doing, develop a different fundamental objection to the meme model. This new objection is that most cultural items are 're-produced' in the sense that they are produced again and again—with, of course, a causal link between all these productions—but are not reproduced in the sense of being copied from one another (see also Origgi and Sperber, forthcoming). Hence

[1] Dawkins adds: 'The difference between high fidelity genes and low fidelity memes is assumed to follow from the fact that genes, but not memes, are digital'. The objection that memes are transmitted with too low fidelity can be made without this further claim, which I find vague and uncompelling.

they are not memes, even when they are close 'copies' of one another (in a loose sense of 'copy', of course).

The objection of low fidelity had been envisaged and taken seriously by Dawkins himself. In *The Extended Phenotype* (Dawkins 1982:112) he wrote:

> The copying process is probably much less precise than in the case of genes: there may be a certain 'mutational' element in every copying event [. . .]. Memes may partially blend with each other in a way that genes do not. New 'mutations' may be 'directed' rather than random with respect to evolutionary trends. [. . .] there may be 'Lamarckian' causal arrows leading from phenotype to replicator, as well as the other way around. These differences may prove sufficient to render the analogy with genetic natural selection worthless or even positively misleading. My own feeling is that its main value may lie not so much in helping us to understand human culture as in sharpening our perception of genetic natural selection.

Of course, what counts as 'too low fidelity' for a given item is relative to the selection bias for that item (see Williams 1966). A greater selection bias allows for a higher mutation rate. On the other hand if, as Dawkins says, there is 'a certain "mutational" element in every copying event' (loc. cit), then it is not easy to see how selection could work at all. It is to this problem that Dawkins (1999) now offers an ingenious solution. He uses for this a thought experiment of which I present a simpler but equally effective version (before discussing his version later). Consider Figure 8.1. A first individual is shown this figure for ten seconds and is asked, ten minutes later to reproduce it as exactly as possible. Then a second individual is shown for ten seconds the figure drawn by the first individual and presented with the same task. This is iterated with, say, nine

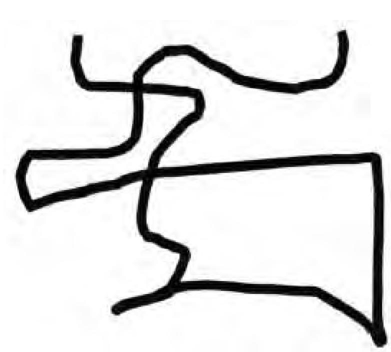

Figure 8.1

participants. It is most likely that each drawing will differ from its model and that the more distant two drawings are in the chain, the more they're likely to differ. A judge given the ten drawings in a random order and asked to put them back in the order in which they were produced should perform, if not perfectly, at least much better than random. The 'mutational elements' in every copying event are such that a drift is manifest, and no stable pattern is maintained.

Now imagine a similar experiment being performed, but this time with Figure 8.2 as initial input. Again, each dewing produced by the successive participants is certain to differ from its model, since each participant will fail to reproduce the model in all its particulars. This time, however, the distance in the chain of two drawings on the one hand, and their degree of difference on the other hand should be two variables independent of one another (or likely so). A judge asked to put the ten drawings in the order in which they were produced should be cable to do better than random. Despite low fidelity of copying, a stable pattern is most likely to endure across versions, and individual variations are very unlikely to compromise this pattern.

What explains the difference between the two experiments? In the case of Figure 8.1, people try and form a mental image of a drawing which they do not recognize in any way, and then try and reproduce this mental image on paper. In storing the information, in recalling it, and in reproducing it, they are likely to introduce unintended variations that are either in random directions, or are in the direction of entropy, that is, plain loss of information. In the case of Figure 8.2, people recognize

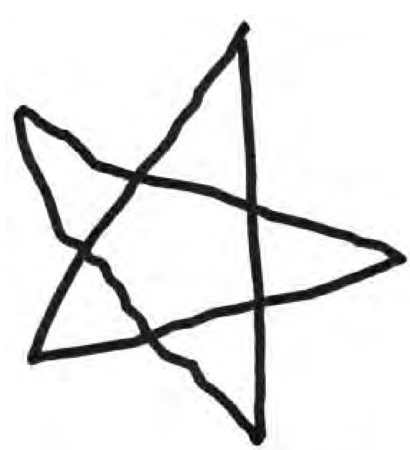

Figure 8.2

the figure as a five-branched star drawn without lifting the pencil. They may well forget most of the other particulars of the drawing under their eyes, such as length of relatively slight segments, or angles. Still, they will produce another star of the same type.

Dawkins might describe the difference between the two types of tasks as follows. In tasks of the first type, what gets copied is the product, the drawing. There is no difference therefore, between the 'phenotype' and the 'genotype', and phenotypic variations are also genotypic variations. In cases of the second type, what gets copied is the implicit instruction ('draw a five-branched star without lifting the pen'). These instructions are the true genotype, while the drawings are only phenotypes. Each participant in the experiment assumes that the preceding participant merely intended to follow the implicit instruction, and that imperfections or idiosyncrasies were unintended and should be ignored.

Individual variations in the productions of the phenotype do not matter. They are not genuine mutations. 'The instructions,' writes Dawkins 'are self-normalising. The code is error-correcting' (1999:xii).

Dawkins concludes the argument by stating: 'I believe that these considerations greatly reduce, and probably remove altogether, the objection that memes are copied with insufficient high fidelity to be compared with genes. For me the quasi-genetic inheritance of language, and of religious and traditional customs, teaches the same lesson' (p. xii). In other words, the stability of cultural patterns is proof that fidelity in copying is high despite individual variations. These variations are phenotypic, not genotypic, and Darwinian selection can take place without being jeopardized by too high a rate of mutation.

I, on the other hand, believe that what is here offered as an explanation is precisely what needs to be explained; what is offered as a solution is in fact the very problem to be solved. Saying that the instructions are 'self-normalising' amounts to resolving a problem by invoking a mystery. The type of thought experiment proposed by Dawkins is well worth analysing so as to solve the mystery. The conclusions I draw from this thought experiment are, however, very different from that of Dawkins. They point to yet another difficulty with the meme model.

Let me grant forthwith two points to Dawkins:

1. Of course, one item A can be a replica (in the relevant sense) of another item B without being identical to B in every respect. From a memetic point of view, it is enough that A and B should share the properties the recurrence of which one is trying to explain.

2. Of course, cultural items exhibit, over periods of time of various length (longer for folktales, shorter for modern dress fashions, for instance), the kind of stability found, on a much smaller scale, in Dawkins thought experiment. That is, although there is much individual variation, items of the same type all remain in the vicinity of one another and instantiate a common pattern.

The issue is whether the relative stability found in cultural transmission is proof of replication. Dawkins seems to think it is. In substance, he proposes a test to decide whether a causal chain that links the production of a series of items is a chain of replications. The test is as follows. Present (or suppose you present) to an intelligent observer the items in the chain in a random order. If the observer finds it impossible to put back, at least approximately, the items in the order in which they were produced, then these items are replications in the relevant sense. Individual variations among these items are phenotypic and do not compromise the stability of the underlying genotype. Much of culture passes this test and is seen, then, as made of replicators.

To show that Dawkins test is not as reliable as it may seem, let me first give an example of a causal chain that would meet the criterion, but could not be properly described as a case of memetic transmission. Consider the case of laughter. Laughter is a social behaviour that is typically triggered, in individual development, by the laughter of others, and that remains a highly contagious form of behaviour. Laughter is influenced in its intensity, style, and circumstances of arousal by cultural factors. Moreover, even within a cultural group, there are important individual variations. Now, imagine a series of registerings of causally linked individual laughters (linked either in the stabilization of laughing behaviour across generations, or in a much shorter causal chain of contagious laughter). If these registerings were presented in a random order, they could not, I take it, be rearranged in their causal order. Laughter passes Dawkins test. Yet, surely, it is not a meme.

Why is laughter not a meme? Because it is not copied. A young child who starts laughing does not replicate the laughters she observes. Rather, there is a biological disposition to laughter that gets activated and fine-tuned through encounters with the laughter of others. Similarly, an individual pushed into convulsive laughter by the laughter of others is not imitating them. The motor program for laughing was already fully present in him, and what the laughter of others does is just activate it.

Let me generalize and define three minimal conditions for true replication. For B to be a replication of A,

(1) B must be caused by A (together with background conditions),

(2) B must be similar in relevant respects to A, and

(3) The process that generates B must obtain the information that makes B similar to A from A.

Another way to express this third condition is to say that B must inherit from A the properties that make it relevantly similar to A. Discussions of memes take implicitly for granted that the co-occurrence of causation and of similarity between cause and effect is sufficient evidence of inheritance. But this is not so. The cause may merely trigger the production of a similar effect, as we saw with the case of laughter. Even if conditions (1) and (2) are satisfied, condition (3) may not be.

Consider a theoretical example, with two cases to be compared. In both cases conditions (1) and (2) are satisfied, but condition (3) is satisfied only in the second case. First case: ten sound-recorders with the same repertoire of melodies in each have been fixed so that they are activated by the sound of the last five bars of any melody in their repertoire, and then play this very melody. They are placed in such a manner and at such a distance of one another that the fist one activates the second, the second the third, etc. The first recorder plays melodies in random order at appropriate time intervals. Second case: ten sound-recorders have been fixed and placed so that the second-recorder records sound from the first, and then replays it, the third recorder records sound from the second and then replays it, and so on. Only the first recorder has a ready repertoire of melodies, and it plays them in random order at appropriate time intervals. In both cases, an observer listening to these devices playing, each in turn, one melody after another, and unable to inspect them otherwise, would have some reasons to think she was witnessing a series of replications. In fact, this would be true in the second case, but not in the just, where only triggering takes place and no copying at all. This illustrates the point that, in the case of a causal chain that satisfies conditions (1) and (2), further evidence about the causal processes involved must be available before one is in a position to argue that condition (3) is also satisfied, and that one is dealing, therefore, with a true chain of replications.

Let us go back, now, to our thought experiment. In the first task (memorizing and reproducing Figure 8.1), participants rely on general perceptual, memory and motor abilities. In other words, they rely on the general human ability imitate, an ability which is taken by memeticists to be extremely powerful. In this case, however, it fails. In the second task (memorizing and reproducing Figure 8.2), the stimulus is recognized. That is, it triggers the activation of pre-existing knowledge. The stimulus is categorized as a token of a general type: a five-branched star drawn without lifting the pencil. Properties of the actual stimulus that are irrelevant to this categorization are just ignored. When asked, after ten minutes, to reproduce the stimulus, participants just produce another token of a five-branched star without, in most cases, even trying to remember what the original figure exactly looked like. Their ability to perform well in this second task is not an ability to perceive and copy. It is an ability to recognize and re-produce, using for this, knowledge of the five-branched star type that they already possessed before encountering the token. It is not, then, that people are better at imitating Figure 8.1 than at imitating Figure 8.2. They are indeed bad at imitating Figure 8.1, and they are not imitating Figure 8.2 but merely producing a new token of the same recognizable type.

Dawkins's original thought experiment involved a comparison of two tasks: reproducing drawing of a Chinese junk, or making an origami Chinese junk after having been taught, by demonstration, how to make one. Unlike my simpler version, the two final products—drawing or the origami—are recognized by the participants. In the drawing version, however, participants are unable to recognize the series of strokes that would yield the full drawing, whereas in the origami version the successive folding are individually demonstrated. Thus, the two task are different, not just in the type of item to be copied (a drawing vs. an origami) but also in the fact that participants observe only the product in the first task, and the process of production in the second task. If participants were just shown a finished origami junk, they would, presumably, do even worse in reproducing it than in reproducing a drawing of a junk.

The crucial difference between the two tasks is that the second involves demonstration, and the other not. From the demonstration, or so Dawkins assumes, participants can and do infer implicit instructions (e.g. 'take a square sheet of paper and fold all four corners exactly into the middle'). These instructions are not a description of what the person making the origami is actually doing (the four corners are never folded exactly into the middle, for instance) but a description of what the person is aiming at, is intending to do. Inferring instruction involves much more than the ability to perceive and describe actual movements; it involves the ability to attribute goals and intentions.

Contrary to what Dawkins writes, the instructions are not 'self-normalizing'. It is the process of attribution of intentions that normalizes the implicit instructions that participants infer from what they observe. When you see the person folding the four corners of a square sheet of paper into four different points in the vicinity of the middle, you assume that she was aiming at the middle rather than at these four odd points Such intentions to realize regular geometrical pattern are familiar—in particular, in the context of origami—and readily attributed. You recognize, in other terms, the behaviour as an imperfect realization of an intention of a familiar and regular type rather than as the perfect realization of an intention of an unfamiliar and irregular type. The instructions that you infer are, then, informed in part by what you actually observe, and in part by what you already know of human intentions, and of the type of instructions typically used in origami.

The instructions are not being 'copied' in any useful sense of the term from one participant to the next. Certainly, instructions cannot be imitated, since only what can be perceived can be imitated. When they are given implicitly, instructions must be inferred. When they are given verbally, instructions must be comprehended, a process that involves a mix of decoding and inference (Sperber and Wilson 1995). The inference involved in either case draws on domain-specific competencies having to do with the attribution of intentions and with knowledge of the role of regular geometric forms in the formation of human intentions generally, and in paper-folding in particular. Thus, the normalization of the instructions results precisely from the fact that something other than copying is taking place. It results from the fact that the information provided by the stimulus is complemented with information already available in the system.

In the real world, and in particular in the cultural world, triggering and copying can and do combine in various degrees. What gets triggered by cultural stimuli are acquisition mechanisms and competencies that are more or less domain-specific. These mechanisms are themselves in part genetically, in part culturally, inherited.

Let us briefly consider the example of the acquisition of language. In acquiring a language, a child internalizes a grammar and a lexicon on the basis of linguistic interactions. Nowhere in these interactions—nowhere in the linguistic data the child is presented with—is the grammar present to be copied. Rather, the grammar must be inferred from these data. As Noam Chomsky has long argued and as has become, if not universally, at least generally accepted today, this requires a genetically determined preparedness to interpret the data in a domain-specific way and to generalize from it to the grammar of the language, going well beyond the information given. Imitation in some sense may well play a role—though not a sufficient one—in the acquisition of the phonology of words, but not in the acquisition of their meaning. Meaning is not something that can be obeyed and copied. It can only be inferred. Language learners converge on similar meaning on the basis of weak evidence provided by words used in an endless diversity of contexts and with various degree of literalness or figurativeness. Acquisition of meaning in such conditions is a feat that would be wholly mysterious if it were not highly constrained by domain-specific competencies having to do with conceptual domains on the one hand, and with the attribution of communicative intentions to speakers on the other. Thus, the similarities between the grammar and lexicons internalized by different members of the same linguistic community owe little to copying and a lot to pre-existing linguistic, communicative, and conceptual evolved dispositions.

The respective role of copying and that of pre-existing dispositions to construe evidence in domain-specific structured ways may vary with different cultural competencies. Learning to tap-dance involves more copying than learning to walk. Learning poetry involves more copying than learning philosophy. For memetics to be a reasonable research programme, it should be the case that copying, and differential success in causing the multiplication of copies, overwhelmingly plays the major role in shaping all or at least most of the contents of culture. Evolved domain-specific psychological dispositions, if there are any, should be at most a relatively minor factor that could be considered part of background conditions. There is nothing obvious about such a view. While the view may have some popularity among unconcerned lay people, no psychologist believes that cultural learning is essentially a matter of imitation (this is true even of psychologists who attribute an important role to imitation, e.g., Meltzoff and Gopnik 1993; Tomasello et al. 1993). In fact, such an idea goes against all major recent developments in developmental psychology and in evolutionary psychology (see Hirschfeld and Gelman 1994). This, together with the problem raised in this article, puts a special burden on memeticists.

Memeticists have to give empirical evidence to support the claim that, in the micro-processes of cultural transmission, elements of culture inherit all or nearly all their relevant properties from other elements of culture that they replicate (i.e. satisfy condition 3 above). If they succeeded in doing so they would have shown that developmental psychologists, evolutionary psychologists, and cognitive anthropologists who argue that acquisition of cultural knowledge and know-how is made possible and partly shaped by evolved domain-specific competencies are missing a much simpler explanation of cultural learning: imitation does it all (or nearly so)! If, as I believe, this is not even remotely the case, what remains of the memetic programme? The idea of a meme is a theoretically interesting one. It may still have, or suggest, some empirical applications. The Darwinian model of selection is illuminating, and in several ways, for thinking about culture. Imitation, even if not ubiquitous, is of course well worth investigating. The grand project of memetics, on the other hand, is misguided.

References

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

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

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

Dawkins, R. (1999). Foreword *The meme machine* by Susan Blackmore. Oxford: Oxford University Press.

Hirschfeld, L. and Gelman, S. (ed.) (1994). *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.

Origgi, G. and Sperber, D. (forthcoming). Evolution, communication, and the proper function of language. In *Evolution and the human mind: Language, modularity, and social cognition* (ed. P. Carruthers. and A. Chamberlain). Cambridge: Cambridge University Press.

Meltzoff, A. and Gopnik, A. (1993). The role of imitation in understanding persons and developing a theory of mind (ed. S. Baron-Cohen *et al.*), *Understanding other minds*. Oxford: Oxford University Press.

Sperber, D. (1985). Anthropology and psychology: towards an epidemiology of representations. *Man* (N.S.), 20:73–89.

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

Sperber, D. and Wilson, D. (1995), *Relevance: Communication and cognition* (2nd edn). Oxford: Blackwell.

Tomasello, M., Kruger, A. and Ratner. H. (1993). Cultural learning. *Behavioral and Brain Sciences* 16:495–552.

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