# The Evolution of Culture

## David Deutsch

A modified version of this previously unpublished March 1994 essay appears in David Deutsch's 2011 book *The Beginning of Infinity*.

***Ideas that survive***

I once heard a former Public-School[1] pupil remark: "the purpose of a Public School is to cause distinctive scars, by which its inmates will be recognisable for the rest of their lives". He was likening the psychological damage that he deemed to have been left by his schooling to the physical mutilations by which many cultures mark their members. This may seem a far-fetched analogy, but it is quite the opposite. It is part of a fundamental truth about the purpose, not just of Public Schools, but of human culture in general. Fortunately it is not the whole truth.

By "culture" I mean any set of shared ideas that make their holders behave alike in some way. Many such ideas are unconsciously held. They include skills, problems, expectations and emotional preferences as well as explicit theories. The shared values of a nation, the shared knowledge of an academic discipline, the admiration of a particular film star, and indeed the after-effects of attending a Public School, are all, in this sense, "ideas" which define cultures. An individual typically belongs to many cultures at once, and to sub-cultures within those cultures.

The major cultures, including nations, languages and religions, social, artistic and philosophical traditions etc., carry information that has been created incrementally over hundreds or even thousands of years. Most of the ideas in these cultures have a long history of being passed from one person to another. In particular, faithful transmission from *older* people to *younger* people is necessary for ideas to outlive human beings.

As people try to solve problems, they sometimes modify cultural ideas in their minds. Inevitably there are also unintentional modifications. So it can only be rarely, if ever, that two minds hold precisely the same idea. Insofar as people transmit a cultural idea to others, they transmit variants of it. Thus a culture must in practice be defined not by a set of strictly identical shared ideas, but by a set of variants that cause slightly different behaviours. If the relative numbers of different variants change, the characteristic behaviours of the culture change accordingly. Some variants, once they appear in one mind, tend to spread throughout the culture. But the overwhelming majority disappear within a generation or less. So the characteristic behaviour of a long-lived culture is determined mainly by ideas that are long-term survivors, and partly by recent variants, most of which are destined to become extinct.

I have intentionally used terms like "survivors", "variants" and "extinct", which have evolutionary connotations. The idea that the theory of evolution applies to human affairs is at least as old as the idea that it applies in biology. However, most attempts to apply it have been based on misunderstandings of evolution. Karl Marx believed that his own theory of history was evolutionary because it spoke of a progression through historical "stages", each of which determined the next according to historical "laws of motion" (which in fact made it decidedly *un*like the real theory of evolution). Nazi ideology likewise misused evolutionary

ideas to justify violence. Although the evolutionary content of Marxism and Nazism is facile, it is no accident that analogies between society and the biosphere tend to be associated with grim visions of society. Biological evolution is a grim process. It involves perpetual war, plunder, exploitation and deceit. Those who think that cultural evolution has similar effects usually end up either opposing it (advocating a static society) or condoning immoral behaviour.

One frequently hears arguments by analogy – comparing a human society to the biosphere for instance, with each person, or each sub-culture, being like a competing biological species; or comparing society to a species, with people as its individual organisms; or comparing society to an organism, with sub-cultures as its organs and people as its cells. There are many variations on these themes, each giving a different conclusion about society, and each supposedly justified by the evolutionary nature of culture. But argument by analogy cannot justify anything without an independent argument for why the analogy should hold. There is no reason to suppose that just because ideas evolve, the "natural history" of cultural evolution must resemble that of biological evolution. As we shall see, although they are described by the same underlying theory, the mechanisms of evolution in the two cases are different, and that makes the outcomes different too.

Richard Dawkins coined the term *meme,* meaning an idea that affects its holders' behaviour and may be passed from one person to another. All systematic resemblances between human affairs and biological processes stem from the genuine close analogy between memes and genes. This analogy holds because both memes and genes are *replicators,* entities that can cause themselves to be copied – or more accurately, replicated – and because in both cases their ability to be replicated depends on causing complex sequences of events in their environment. Information about how to cause those events is encoded in the structure of the memes and genes. In memes such information is called *knowledge,* and in genes, *adaptation.* An entity is *adapted* to causing certain events if it causes them more reliably than all, or nearly all, its close variants do. Biological adaptation is in many ways analogous to human knowledge. But to draw conclusions about culture from the premise that ideas evolve, one should not argue by analogy. Instead one should apply the theory of evolution *directly to memes*. Sometimes the results are reminiscent of biology. More often, despite the underlying analogy, they are not.

The central problem of biology is that of the origin, both in general and in particular cases, of the complex adaptations that species possess. That is, we want to understand the mechanisms by which adaptations come to be encoded in genes, and we want to explain the origins of particular genes. Similarly, we want to understand the general mechanisms by which cultures acquire their complex attributes, and we want to explain the presence of particular memes, such as those responsible for the phenomenon of the British Public School.

It all comes down to the following question: what property distinguishes the successful variant of a gene or meme from its many unsuccessful rivals? The general answer has been given by Dawkins. The successful variant changes the behaviour of its holder in such a way as to make that variant more likely than its rivals to be transmitted to younger holders.

This is one of those deceptively deep truths that is commonly criticised either for being too obvious to be worth stating, or for being false. The reason, I think, is that while it is self-evident that a population of replicators subject to variation will be taken over by the variants that are most effective at getting themselves transmitted, this is often counter-intuitive as an explanation of the adaptations that the winners embody. Although it *is* the explanation, we

are, as Dawkins puts it, "obsessed with purpose", and we intuitively try to explain complex functionality by seeking its purpose, i.e. the objective to which it has been adapted. So if an animal sacrifices its life to protect its young, and we see that this behaviour does not promote the animal's own welfare but does benefit the species, it is tempting to reach for explanations such as "the genes for such behaviour have been adapted to benefit the species". But that is not so. For one thing, what "benefit" of the species do we mean? An increase in the number of individuals? Or an increase in the time for which the species survives before extinction? Or an extension of the species' geographical range or the number of niches it occupies, or could occupy – or what? In general, each of these "benefits" would be maximised by a different variant of the species' genes. But in any case, suppose that some variant of a gene is beneficial in every one of these ways, and let it benefit each *member* of the species as well for good measure, but let it only be poorer than one of its rivals at getting itself replicated, and it will certainly disappear from the species' gene pool. Conversely, if a variant tends to be replicated at a higher rate than the other variants that are present, there is an end of the matter. It does not then make the slightest difference whether the *species* will thereby be benefited, harmed, spread throughout the world or reduced to extinction: that variant will displace the others. So it is not the welfare of species or individuals to which evolution adapts genes, but only their own replication.

How can a gene be bad for the species if it is good at getting itself replicated? Because what counts is the rate of replication of a gene *relative to its rivals*. Suppose that the total number of birds of a particular species would be maximised if they all nested in (say) April, because of a greater food supply, higher temperature or whatever. It might nevertheless be true that a March-nesting variant would take all the best nesting sites. Therefore as soon as a March-nesting variant arose, it would spread through the population. This would reduce the number of birds, so from the point of view of the species and its members, nesting in March is a dangerous "mistake", caused by the "selfishness" of their genes. But evolution does not care. It can even favour genes that are wholly harmful to the species. For example, the peacock's huge tail makes it harder for peacocks to evade predators, and diminishes the viability of the peacock species. During the evolution of the peacock, the reason why genes for larger tails spread through the population was not that they conferred any benefit on the species, but simply that peahens tended to choose large-tailed males as mates. This preference was itself caused by a gene which co-evolved with the males' large-tail gene, but the detailed mechanism does not concern us here. What matters is that on balance, large-tail genes were better at getting themselves replicated than small-tail genes, and so came to predominate. The species and the individuals just had to suffer the consequences.

The length at which the peacock's tail stabilised was determined by a balance between factors that made longer- or shorter-tail genes more likely to be replicated (such as peahens' preferences, and vulnerability to predators respectively). The nesting behaviour of our hypothetical birds is determined by a similar balance. A February-nesting variant would have an even greater advantage in getting the best sites than the March-nesting one. But it would also have greater disadvantages, and when its holders' offspring died of cold or starvation their better-placed nests would do the gene no net good. Evolutionary stability would be reached only when these two effects, and many others, exactly balanced. The gene variant corresponding to the balance point would be producing, on average, more surviving copies per gene than any rival variant, i.e. it would be optimally adapted for replication. If we observed such a species, we could expect to see few if any individuals carrying the April-nesting variant, the one that would make the *species* optimally viable.

The evolutionary balance is not between what is best for the species and what is best for the gene. It is only between different factors affecting the replication of rival gene variants. If the best-replicating variant confers sufficiently large disadvantages on the species, the species must become extinct, and nothing in biological evolution can prevent that. This has presumably happened many times in the history of life on Earth, to species less lucky than the peacock. It follows that all the surviving species in the biosphere have genes that are less harmful than that. Theoretically this effect, known as *species selection,* could create complex adaptations for the species' survival. But extinctions of species (or of differently adapted, non-interbreeding sub-species) do not happen very often – once every few generations at most. Moreover, species selection is choosing from a very small pool of variants, each variant being an entire population that does not interbreed with other populations. (Incidentally, the total number of species that have ever existed is thought to be only about a thousand times larger than the number that exist today.) In any case, after a species-selection event, the gene variants in the surviving sub-species are worse replicators than certain other variants – namely those which caused the extinction of the extinct sub-species. So if the latter variants are not to arise again, and undo the nascent adaptation that benefits the species, there must be further evolution that favours the *replication* of the surviving variant gene. Thus a lengthy and unlikely sequence of events is required for a species to benefit even once from species selection. A chain of many such selections would be required to create a significant adaptation. Therefore, species selection notwithstanding, we cannot expect to see genes that are adapted to benefit the species.

Is it a coincidence, then, that most genes do confer some, albeit less than optimal, benefits on their species, and on their individual holders? No. Organisms are the slaves that genes use to achieve their "purpose" of replicating themselves. Genes make a living, and gain competitive advantages over each other, just as human slave owners did: in part by keeping their slaves in good working condition. An owner might well sacrifice individual slaves to ensure the efficient operation of that owner's slave workforce, just as genes sacrifice individuals for the species' benefit. But slave owners did not set up their businesses and households for the purpose of keeping their slave workforces in existence, any more than they did it for the benefit of individual slaves. They fed and housed their slaves, and indeed forced them to reproduce, so that the slaves could work to achieve the *owners'* objectives. So it is not surprising that genes usually confer some benefits on their species and its members. Nor should it surprise us that they often harm them. But what genes are *adapted* to – i.e. what they do better than almost any variant of themselves – has nothing to do with the species. It is getting themselves replicated. What explains the presence of one gene variant rather than another is its relative excellence in replicating itself in the environment in which it evolved, and that alone.

***The selfish meme***

All this holds for memes too. Memes, like genes, are "selfish". That is, what makes one variant of a meme spread while others die out is not that it benefits its holders, or their society, but only that the successful variant changes the behaviour of its holders in such a way as to make that variant more likely than its rivals to be passed on to younger holders. The analogue of species selection, namely the selection of surviving *cultures,* is even less significant in meme evolution than in gene evolution, because it is quite rare for an idea that has spread within some culture, and is capable of spreading in others, to become extinct even if the original culture is destroyed. Therefore the explanation of the presence of a meme in a culture is never that rival cultures with variants of the meme have become extinct. It is always that rival memes have become extinct within that culture. And they became extinct because they were

less good than the survivor at getting themselves replicated within that culture. The spread of the best-replicating variant may benefit its holders, or society as a whole. If it does not benefit them, it will happen anyway.

Fortunately, in the case of memes, that is not the whole story. As I have said, despite the underlying analogy between memes and genes, they evolve by different mechanisms. The most important differences are these: while variations in genes are purely random, memes are also subject to conscious variation, intended to achieve desired effects; and while genes are rejected only if their host fails to reproduce, memes can be rejected intentionally *by* their host. A human mind has purposes of its own and is capable of modifying memes in its efforts to achieve those purposes. Moreover, a mind contains ideas which it has created itself, ideas which have never existed, and will never exist, in any other mind. So such ideas are not memes. Yet some of them are preferences, and purposes, and will be used to choose which memes the person will act upon. Hence the "selfish" purposes of memes are not the only ones that can affect the course of meme evolution. This is where the analysis of meme evolution begins to differ markedly from that which any direct analogy with genes would suggest.

Because of the operation of these two independent sets of purposes – those of the meme and those of its holders – a meme has to run the gauntlet of two independent selection processes alternately during its life-cycle. It must first, like a gene, code for some behaviour which causes it to be transmitted faithfully into other minds. The simplest example of such behaviour is teaching ideas about parenting to a child, which is analogous to copying the genes for reproductive organs into an offspring's genome. When it has caused such copying, a gene has effectively completed its life-cycle. The offspring are guaranteed, given the opportunity, to re-enact the behaviour and copy the gene to *their* offspring's genomes. By contrast, a meme, at this stage, has barely begun its life-cycle. Merely creating a copy of itself in a young person's brain will by no means ensure that that person will in due course enact the corresponding behaviour. For a human brain, unlike a genome, is itself an arena of intense competition between multiple variants of memes. Indeed, few ideas within a brain are direct imitations of outside ideas. Most are variants deliberately created by the brain for the very purpose of subjecting them to selection – and then to further variation and selection which can easily run to thousands of evolutionary cycles before the survivor, perhaps transformed beyond recognition, is ever acted upon. Moreover, this second type of selection not only uses different criteria from those of biological evolution, it uses a qualitatively different method. Unlike biological evolutionary systems, human minds are capable of knowing what ideas mean. Biological evolution can only test an idea by trying it out in practice, and failure involves the death or incapacitation of the holder. But we can *criticise* ideas. We can examine their meaning and decide, without having to enact them, whether we think that doing so would further our purposes or not. For every idea that we test in practice, we first test thousands in our imagination. And we can vary each meme independently, without having to throw away the whole collection whenever a variant of a single meme turns out to be non-viable. As Popper puts it, we "let our ideas die in our place". Thus the overwhelming majority of memes that fail to be propagated, fail not because their holder dies or is physically prevented from propagating them, but because the holder chooses not to act upon them and acts upon variants instead. That is something that cannot happen to a gene (except in a biotechnology laboratory).

Another difference between memes and genes is that memes are not replicated by being physically copied, as genes are. We do not have access to the physical representations of ideas in people's brains (and in any case, it is unlikely that different brains use the same internal "code" to represent ideas). Instead, we have to work our ideas out for ourselves, using other

people's behaviour as a clue. Among the first ideas that we have to acquire are those that define language. Most memes are expressed partly in language; nevertheless every significant facet of culture has an inexplicit component as well. One has to interact with existing members of the culture, observing their behaviour as well as remembering what they say. This indirect form of replication has a number of evolutionarily significant consequences. The most obvious one is that it is more open to the accumulation of what from the memes' point of view are errors, but from the individual's point of view are opportunities to *correct* errors, i.e. to adapt the cultural knowledge according to that individual's preferences. Another consequence is that there is no predetermined "inheritance" structure for memes. A gene is passed from parent to offspring and descendants with very high fidelity, but cannot (in nature) be passed at all to anyone else. Memes can be transmitted from anyone to anyone (in this respect they are more like viruses than genes), and on a time scale far shorter than a human generation, but there is no automatic way of transmitting them reliably.

All this makes the evolution of memes intrinsically much faster and more flexible than that of genes. During one human generation, which allows only one cycle of variation and selection of genes, there may be thousands of generations of meme evolution. Hence the frequently cited metaphor of the history of life on Earth, in which human civilisation occupies only the final second of the "day" during which life has so far existed, is based on a misunderstanding of the nature of evolution. A substantial proportion of all evolution to date, perhaps even most of it, has occurred *since* the arrival of our species. It has occurred in our brains. The whole of biological evolution was but a preamble to the main story of evolution on our planet, the evolution of memes.

***Static societies***

Given that a meme which is to get itself transmitted for many human generations must somehow repeatedly survive intense competition under two independent types of variation and selection, one might consider it a miracle that any meme manages to be transmitted more than once without being changed. Nevertheless there exist memes that achieve this miracle repeatedly. Let us consider the survival strategies of these longest-lived memes, the memes responsible for such things as nations, societies and religions.

Our society – the social and political culture in which we in the Western world have been living for the last few centuries – is unique in that it is perceptibly changing. Historically, most long-lived societies have been almost completely static, and none but ours has ever changed rapidly enough for its members to notice. Of course *some* things have always been subject to rapid change: famines, plagues and wars have occurred; kings have died and new kings have been crowned. But that did not affect the memes governing behaviour in any important aspect of life. People could expect to die under much the same moral values, personal life-styles, technology and pattern of economic production as they were born under. Such things changed noticeably only at moments when one society was violently superseded by another. While a society lasted it was, from its members' points of view, static.

Before we can understand our unusual, dynamic sort of society, we must understand the usual, static sort.

What does it take to make a society static? It takes fidelity in the replication of the relevant memes. If the society is to remain static for many generations, then despite the inherent unreliability of meme transmission, very high fidelity is required. It has to be even higher for memes than for genes because a meme variant can be designed, rather than randomly

mutated, and so might cause a much larger complex change in behaviour than a gene variant could. Moreover it is impossible to tell in advance which variants have the capacity to spread, so from the point of view of a static culture, all variants of existing memes are dangerous (i.e. they threaten to change the culture). The most straightforward way of increasing the fidelity of transmission is to suppress the expression of variant memes, i.e. to suppress dissent and deviation from cultural norms of behaviour. Thus every static culture has its own version of a secret police, or an Inquisition, or a headmaster, whose task is to prevent change in the culture's constitutive memes. However, it is in the nature of information in general, and especially of any idea that some people find preferable to prevailing ideas, that suppressing it is very difficult and expensive, and fully eradicating it is almost impossible. No culture could remain static solely by preventing people from transmitting and acting upon dissident ideas *once they had been created.* That is why the enforcement of conventional behaviour is only ever a secondary, mopping-up operation to catch the few variant memes that have evaded the primary, very efficient method by which static societies prevent change. That method is systematically to disable the *source* of new ideas, human creativity. The main targets of this are always children, for two reasons. First, if one waits until potential dissidents have grown up, it may be too late: they may already have conceived and begun to spread their heretical ideas. And second, childhood is the easiest and most effective time to apply the sort of treatment to people that disables their creativity. What sort of treatment is that? My current argument does not depend on the answer to that question, but let me say parenthetically that I believe that what disables children's creativity is *coercion*: making them do things, or refrain from doing things, against their will. Coercion itself does the job, rather than any specific script that children may be forced to enact, though coercion in a particular area of life tends especially to impair a child's ability to think creatively in that area.

Keeping the society static is not, however, the purpose to which any individual static-society meme is adapted. Its purpose is to replicate itself, i.e. to cause behaviour towards children of precisely such a kind as to make them, in due course, behave in the same way towards *their* children (or pupils, etc.). Therefore the loss of creativity suffered by the recipient of such behaviour cannot be indiscriminate. On the contrary, it must take the form of a finely tuned psychological disorder: the inability not to re-enact the behaviour that caused it. The most straightforward examples of such memes are found in religions. Long-lived religions invariably prescribe distinctive child-rearing behaviours, whose purpose is to induct children into the religion by causing such states as faith, guilt, and fear of the supernatural. This psychological impairment is deeply and subtly entrenched in the recipients' minds, so that they find themselves facing a large emotional cost if they subsequently attempt to deviate from the religion's prescribed behaviours. Thus we have the common but apparently perverse phenomenon of people continuing to enact religious rituals after they have ceased to believe the doctrines that ostensibly justify them. Predictably, the cost of deviation is greatest when it comes to behaviour towards one's children. Many parents who no longer have religious beliefs nevertheless submit their children to religious education, so that the practice of doing so can survive for another generation.

How do the memes in question know how to achieve such complex, reproducible effects on the behaviour of human beings? They do not, of course, know how to, in the sense of having explicit theories or intentions. It is just that they exist, at any instant, in many variants. For every meme that forms part of a static culture, millions of variants will have fallen by the wayside precisely because they lacked what the survivor had: that extra piece of information, that extra degree of ruthless efficiency in preventing variants from being acted upon, that slight advantage in psychological leverage, or even actual benefits that it provided – whatever

it took to make it spread it through the population in competition with its rivals – and, once it was prevalent, to get it replicated with just that extra degree of fidelity. If ever a variant happened to be just a little better at inducing behaviour with those self-replicating properties, it soon displaced all other variants. As soon as it did, there were millions of variants of that variant, which were again subject to the same evolutionary pressure. Thus, in successive generations, successive versions of the meme accumulated the knowledge that enabled them ever more reliably to inflict their characteristic style of damage on their human victims. Just as the genes determining the structure of the eye seem to know the laws of optics, long-lived memes may seem to possess superhuman wisdom about the human condition, and to use it mercilessly to evade the defences and exploit the weaknesses of the human minds that they enslave.

Because static societies survive by effectively eliminating the type of evolution that is peculiar to memes (namely variation by means of the holder's creativity, and selection according to the holder's preferences), they resemble biological systems more closely than cultures in general do, and have some of the unpleasant features that naïve evolutionary analyses assume are inevitable in general. Yet according to Rousseau's theory of the "noble savage", individuals in primitive societies were free from the restraints of social convention, free to achieve self-expression and fulfilment of their "natural" needs, while at the same time avoiding the alienation from others that was supposedly produced by the competitive and acquisitive imperatives of civilisation. Other critics of our society have attributed every fashionable moral value – from military virtue to family values to sexual liberation to ecological awareness – to primitive societies of the past or present. Be that as it may, insofar as primitive societies were and are *static* (and if ever one ceased to be static it would soon cease to be primitive, or else destroy itself) they cannot afford to allow their members much opportunity to pursue happiness. Any time or effort that is not devoted, directly or indirectly, to the faithful propagation of memes is, from the memes' point of view, wasted. Moreover, the pursuit of happiness cannot get far without the exercise of creativity, and creativity risks change. Consider the standard fantasy of a primitive society in which people happily live out their traditional roles for generation after generation. They may have no complaint about those roles because they are emotionally and practically committed to them, and in any case can imagine no other way of life. Nevertheless they are not immune from pain, hunger, grief, fear or other forms of physical and mental suffering. Imagine that someone, somewhere has a new idea for alleviating some of that suffering. It need only be a once-in-a-lifetime inspiration in the mind of an otherwise nondescript, conforming citizen. It need only be a small, tentative improvement: a way of growing food with slightly less effort, or of making slightly better tools; a subtle change in the relationship between husband and wife, or between parent and child; a slightly different attitude towards the society's rulers or gods. If the change seems to make life a little better, why will the originator not tell other people about it? And if those people agree that the idea improves their lives, why will they not start telling their families and friends, and they theirs? If that happens, the new way of thinking or doing things will soon be the prevailing one. That change will in turn have consequences, creating new problems and opening new opportunities for other people to have ideas that will improve *their* lives. Multiply that by the number of minds in the population, and exponentiate it over a few generations, and it becomes a revolutionary force transforming every aspect of the society. Yet in a static society, none of this ever happens. Why not? It must be because, contrary to what we have just supposed, no such idea is thought of in the first place. Thousands or millions of human beings, wishing throughout lifetimes and generations for their suffering to be alleviated and their lives to be improved, fail to make progress in realising these wishes, and largely fail even to think about ways of doing so. The creativity-suppressing

mechanisms that cause this do "benefit" the society, in the sense of perpetuating it (though they make it more vulnerable to the external forces that will eventually destroy it); they certainly "benefit" themselves, by creating an environment in which they are faithfully replicated. But from the point of view of every individual in the society, those mechanisms are harmful. Thus a static society, far from leaving its members free to achieve self-expression or fulfilment, must leave them chronically baulked in their attempts to address life's problems. It cannot afford to do otherwise. It can perpetuate itself only by thwarting their aspirations and breaking their spirits, and its memes are exquisitely adapted to doing so.

The constitutive memes of a static society are as selfish – as indifferent to human welfare – as any gene, and can be far more sophisticated and efficient in achieving their purposes. Since they ensure their own faithful replication by disabling the thought processes of children in reproducible ways, a static society can be regarded as just a distinctive and stable system for harming children. So if ours were a static society, that characterisation of Public Schools – and schooling in general – would have been correct, but too lenient. Insofar as a school can be understood in static-society terms, the "scars" that make its former inmates *recognisable* are a mere flourish on top of more deep-seated psychological damage. Inflicting that damage is (in a static society) the core function, indeed in a sense the only function, not only of all schools but of all child-rearing and of the society as a whole.

***Dynamic societies***

But ours is not a static society. It is the only known instance of a long-lived dynamic society. The constituent memes of such a society, though still selfish, are not necessarily harmful to individuals. That is because a rapidly changing culture allows the emergence of an entirely different class of memes which, if the society can continue to evolve for long enough, must eventually displace all the anti-humane memes of a static society.

To explain the nature of these new memes, let me pose the question, what sort of meme is it that can cause itself to be replicated for long periods in a rapidly changing environment? Does it even make sense to speak of a meme being "replicated" given that the behaviour that the meme causes is continually changing because its effect depends on many other memes which are themselves changing? The problem presented to a meme by a changing environment is just an extension of the problem I have already mentioned, of being subject to variation and selection according to diverse, unpredictable criteria chosen by individual holders of the meme. The memes of a static society get round the problem by conspiring to eliminate that type of variation and selection. In a dynamic society, that conspiracy has by definition failed, because the society is perceptibly changing on a time-scale shorter than that on which the remaining form of meme evolution (based on imperceptible "errors" in transmission from one generation to the next) can respond.

That is not to say that every static-society meme immediately ceases to be viable once the society begins to change. Such a meme will still be replicated so long as it can manage to cause the special psychological effects that prevent its holders from abandoning it. However, once society is changing, there is always a risk that the knowledge in the meme, which is adapted for replication in the original environment, may be insufficient to cope with some newly arising selection pressure. Because it would only be capable of evolving slowly, it is unlikely that a variant of the meme with enough knowledge to survive could evolve in time. So an entire class of meme variants causing similar behaviours would become extinct.

So I return to my question, what sort of meme – what sort of *idea* is it, that tends to survive criticism from diverse, unpredictable sources? The answer is, a *true* idea. And not just any truth will do. It has to be a useful truth, so that it will have a chance of meeting people's preferences. Useful to whom? To individual people, for it is they who will be choosing whether to enact it or not. "Useful" in this context does not refer only to practical utility, but to any property that can make people want to adopt an idea and pass it on to others, such as being interesting, beautiful, easily remembered, morally right etc. What this really amounts to is that it has to be a *deep* truth, so that it has a chance of retaining its utility when preferences change. Of course, people are fallible and have many preferences for false, shallow, useless, or morally wrong ideas; but *which* false ideas they prefer will differ from one person to another and will change with time, whereas a true, deep idea has a chance of being considered useful by diverse people with diverse purposes over long periods, and therefore has a chance of becoming a meme in a dynamic society.

Because memes of this new type survive by harnessing their holders' critical thought processes, let me call them *rational* memes; and let me call the older type, which survive by disabling their holders' critical faculties, *anti-rational* memes.

Just as anti-rational memes are at a disadvantage in a dynamic society, a static society is unlikely to contain many rational memes. That is because the replication strategy of a rational meme requires its holders to understand that it is a useful idea, so that they will choose to enact it in preference to its variants. This gives them the incentive and the means to apply the idea to new problems, and to seek further ideas along the same lines. The truer and deeper the idea is, the more likely they are to succeed, and every time they do, they change society. For example, a true scientific theory may lead to the invention of new technology which changes the society's traditional economic arrangements, or new weapons which change the traditional balance of power, and so on.

Presumably *some* of the ideas in a static society do survive unchanged just because people prefer them the way they are. But there cannot be too many of those. For even the most innocuous-looking, straightforward, factual truth, if propagated rationally, can easily become a source of unpredictable changes. Take the idea that a certain fruit is poisonous. People who believe this idea will realise that it could save lives, and will want to remember it and to tell other people so that it can be put to that use. But because successive holders understand the idea, some of them may invent new uses for it. They may realise what would happen if their enemies ate some of the fruit, and so the art of poisoning would take its place among the society's memes. Other people may discover that small doses of the poison cause pleasant sensations. Still others may find medical uses for it, and so on ad infinitum. The discovery of each new use constitutes a change in the society, and that change may cause further changes. Conversely, if a society has been effectively static for generations it cannot have tolerated many rational memes. By contrast, memes expressing false ideas, propagated by anti-rational methods, are unlikely to create any such instability. It is safe, for instance, to let people believe firmly in the power of the Holy Spirit. No one will ever succeed in harnessing it in an unexpected way, such as bottling it, thereby revolutionising the church and society. Nor is it necessary for anti-rational memes to be useful to, or understood by, their holders. In fact the more they harm their holders – provided that the harm is of the special, highly adapted type that will replicate the meme – and the less the holders understand them, the better they work.

So although we can imagine memes propagating rationally in a static society, that is not the general way of things. A long-lived static society must transmit almost all its truths, even very useful ones, by anti-rational means which conceal, rather than rely on, the reasons why they

are useful. For example, people's eating habits in static societies tend to be controlled by rituals and taboos, not theories of nutrition and hygiene. Such memes preserve some useful truths, but at the cost of preventing them from being improved upon, and equally they preserve other ideas which greatly harm, indeed which survive precisely because they harm, their holders. It may seem reasonable that people would tend to pass on ideas just because other people would want to know them. But that is not how people behave in a static society. Instead they feel obliged to communicate with one another through stereotyped patterns of meme-based behaviour. For example, until recently there was no sex education, even though parents knew that their children might wish to know the truth about sex. But in the static precursors of our society, there was a meme that forbade the direct transmission of these truths, and instead caused adults to lie to children about sex. Moreover they did this in such a way as to engender shame, guilt and insecurity of precisely the type that would eventually compel the children to tell the same lies to their children.

In summary, there are two types of meme, with nearly opposite replication strategies. Anti-rational memes evolve and survive best in a static society, and jointly prevent change in that society. Rational memes evolve and survive best in a changing society, and they promote further change. Each type of meme can survive in the "wrong" sort of society, but it is at a disadvantage there. Both types of meme are selfish, in the sense that they are adapted for replicating themselves rather than for benefiting or harming society. Both types often benefit society, though not optimally, and sometimes harm it. Considered as an idea in people's minds, an anti-rational meme tends to be a false idea, and needs to harm its holders in order to replicate itself. A long-lived rational meme tends to express deep truths, and needs to benefit its holders in order to replicate itself.

The chief thing that the two types of meme have in common is that they embody *knowledge:* factual knowledge about the outside world, and knowledge of psychology, morality, language, aesthetics, and so on. This may seem odd, since I have just explained why an anti-rational meme is likely to be a false idea. The contrast here is between the knowledge that a meme may constitute as an idea in its holder's mind (which for an anti-rational meme is likely to be false), and the knowledge that it embodies implicitly, of which the holder has little inkling, and which for any long-lived meme is likely to include a lot of truth. Consider religions again. The memes of a religion are replicators because they cause each generation of believers to enact behaviours which cause the next generation to enact them too. No one knows exactly what it is about those memes that gives them this rare self-replicating property. In particular, the believers themselves do not know. Often they fear that *any* deviation from their traditions would risk the religion's extinction. (They are probably mistaken, because the religion could not have spread in the first place without being stable – but their exaggerated fear is part of the meme's strategy for stabilising the religion.) When a religion is in decline, its members typically have no idea how to put things right, because they have no idea how it worked when it was working successfully. Nor can anyone predict which of two religions is just a fad and which will still be there a century later. Yet a successful religion survives because it succeeds in tapping into some deep truths about human minds. In that sense it may be said to embody knowledge of those truths, which it uses to promote its replication. Those truths may benefit the meme's holders, or society, or they may not. But what the meme's *holders* know, or rather believe, is unlikely to be true, or to benefit anyone. If a certain type of hobgoblin has the property that, if children fear it, they will grow up to make *their* children fear it, then that type of hobgoblin (or rather, the behaviour of telling stories about it) is a replicator. The more accurately the hobgoblin's attributes reflect genuine, widespread vulnerabilities of the human mind, the better able it

will be to replicate itself. If the meme is to survive for many generations it is essential that its knowledge of these vulnerabilities be true and deep. But it is far from essential that the content of the meme *qua* idea – that is, the idea of the hobgoblin's existence – have any truth in it at all. On the contrary, the very non-existence of the hobgoblin is likely to make the meme a better replicator (because, for instance, the storyteller is then unconstrained by the mundane attributes of any genuine menace, which are always finite and to some degree combatable).

***Abstract replicators***

Let us consider more precisely what memes are. I have said that a replicator is an entity that causes itself to be copied, in a given class of environments. In doing so, it typically causes many physical objects and processes to appear, at given stages in its life-cycle, each of them helping to cause the appearance of the next. For example, the peacock's tail grows during the peacock's youth, then plays its role in the mating ritual, which in turn leads to the laying of an egg containing a DNA molecule, which later causes the appearance of another tail, and so on. So are all these things – the DNA molecule, the tail, the mating ritual and the egg – replicators? No. For among all these periodically appearing objects and phenomena, only one is actually *copied,* and that is the DNA molecule. Nothing copies the tail, for instance. If we dye it red, we may decrease or increase the peacock's chances of reproducing, but if it does have offspring, they will certainly not have red tails. Their tails will have patterns determined, as usual, by the genes in their parents' DNA molecules. So the tail is not a replicator. But suppose we use biotechnology to alter the DNA molecule in the peahen's egg. We change a base-pair in a gene that codes for the tail pattern. Again, we may decrease or increase that egg's chances of producing offspring. But all the male offspring it does produce will have differently patterned tails, and so will *their* offspring, because our DNA molecule will be *copied* when the offspring reproduce. So the DNA molecule, and it alone, is the replicator responsible for all the periodic phenomena in the peacock's life-cycle.

Similarly we can ask which of the objects or processes involved in cultural evolution are replicators. As I have said, the replication of memes never involves any direct copying, and a meme may well have a different physical form when represented in the brain of each of its holders. Nevertheless, we can apply the same criterion to locate the replicator as we do with genes. Consider anti-rational memes first. A religion, for instance, is associated with a variety of periodic phenomena. Physical representations of ideas in the brains of believers cause behaviours such as prayer, religious education of children, etc., which eventually cause a new generation of believers to hold the same ideas. When religious ideas are accidentally varied, a variant form is sometimes incorporated into the beliefs of subsequent generations. Dawkins' favourite example is the Christian doctrine of the virgin birth, which is thought to have begun as a translation error – a variant idea in someone's mind. It follows that religious ideas in people's minds are replicators. This is much as we should expect by analogy with genes, except that it is only the meanings of memes, and not their physical forms, that remain constant. But there is more to it than that, for there are physical phenomena at other stages of a meme's life-cycle that *are* replicators. The recipient of a meme receives it entirely by observing the old holders' behaviour. So suppose we were to alter that behaviour slightly, without altering the old holders' beliefs (say, by means of threats). If the altered behaviour retained the ability to instil religion into the next generation, it could be a variant religion that was instilled. For example, if all the adherents of some religion were to paint their children's foreheads red, and tell them that this is a religious duty, then if the children adopted the religion they would paint *their* children in the same way. Their descendants would then inherit this forehead colour in the way that a peacock can *not* inherit its parent's tail colour. This shows that religious or other meme-based *behaviours* (rather than just the ideas in

people's minds) can be replicators. Gene-based behaviours are never replicators (though some animals do have a rudimentary culture, i.e. memes, and such animals do exhibit behaviours that qualify as replicators).

Not every form of behaviour caused by memes is a replicator. The content of silent prayers, for instance, may be a consistent part of the characteristic behaviour of a particular religion, but it is not a replicator because it is not copied. Replicators are to be found only among those meme-induced behaviours that somehow impinge on the next generation of believers, usually children.

It seems that for anti-rational memes we can regard two quite different entities, abstract ideas and physical behaviours, as the replicators corresponding to the meme. Quite generally, underlying every physical replicator there is an abstract one, the replicator's *knowledge*. Considering knowledge as a replicator becomes unavoidable in the case of rational memes, to which I now turn.

There too, the brain representation of a meme alternates with a behavioural one. But in a dynamic society, it is unlikely that either of these physical representations could be a long-lived replicator. Take, for example, Newtonian mechanics, one of our society's oldest rational memes. Newton originally expressed it in Latin, and in an idiosyncratic mathematical notation. Each generation of physicists and mathematicians since then has expressed it not only in different languages and notations (which would be relatively trivial – though note that *religions* sometimes insist on retaining their original languages, just in case!), but in successive conceptual frameworks that were very different from Newton's and from each other – for instance Laplace's concept of the gravitational potential. In modern times, Newton's theory has been superseded as a fundamental description of nature, and is retained only as a useful approximation for making predictions. Moreover, it is not only the concepts of Newtonian mechanics that have changed. The many behaviours that it has caused or affected have been changing too. Most significantly, the behaviours that transmit the meme – who is taught it, and at what age, and in what order, and what explanations they are given, and what examples they apply it to for practice – have all differed from each generation to the next, and nowadays, from each decade to the next. There has been no faithfully repeated behaviour – no equivalent of a religious ritual – associated with the transmission of the Newtonian mechanics meme.

If neither the physical representation of Newtonian mechanics in people's brains, nor the concepts in people's minds, nor the behaviour that the meme has caused, have ever been repeated for long, what has? Only the knowledge that the meme expresses. That has been preserved while all the physical and mental phenomena that have been preserving it have themselves kept changing. The reason why Newtonian mechanics has been useful in such a wide and unpredictable range of applications, and as a starting point for so many other discoveries, and why, as a replicating idea, it is if anything more vigorous today than ever, is that it contains deep truth. But the way we conceptualise that truth, and the way we express it in words and symbols, and the way we use it, keep changing in the light of new discoveries. Only the abstract truth itself, both the implicit kind that is unknown to the holder, and the explicit content that meets the holder's preferences, has passed unchanged from Newton to the scientific culture of the present day.

***The Enlightenment***

A long-lived dynamic society seeks and preserves knowledge that is useful to its members. A static society is also knowledge-preserving, but not knowledge-seeking since its constitutive memes prevent knowledge-seeking behaviour. A static society, in its imperceptibly slow evolution towards more faithfully-replicating memes, effectively seeks only more staticity, which is often called in this context "certainty", "security", "sustainability" or "stability". We may define a *stable* culture as one which tends to preserve its knowledge when challenged by small, unexpected problems. Thus a long-lived static society is presumably stable. But a dynamic society that has long-lived memes is also stable. Though the forms of its memes are continually being modified, one can see with hindsight that this process creates an ever larger store of knowledge, which is preserved not because there is any mechanism preventing people from changing it, but because people find it useful.

Presumably no completely static society has ever existed. Societies have, as I said, changed more slowly than their members could detect, but that still allows for change on a time-scale longer than a human lifetime. The evolutionary tendency towards more faithful replication is unlikely to prevent imperceptibly slow changes, for memes cannot act to reverse a change that is imperceptible to their holders. Thus static societies continue slowly to evolve even after they have become "static". Furthermore, no society exists in a strictly static external environment, nor can new types or combinations of events within the society ever be entirely anticipated. Therefore every society faces a stream of unpredictable problems which would, if left unsolved, increasingly impede the replication of the society's memes. Sometimes problems arise rapidly, within one lifetime. All long-lived static societies have used the same strategy for dealing with such problems: they treat a certain small class of people differently from the rest, in a way that allows them to retain and use a modicum of creativity. From these privileged people (usually the "ruling class" but sometimes clerics, academics et al.), the society and its members get the benefit of having some of that creativity applied to the society's problems. The controlling memes presumably benefit them because a creative privileged person is in a better position to entrench that form of privilege than a non-creative privileged person. But this strategy also incurs the constant risk that new, destabilising ideas might be inadvertently created. Indeed virtually all successful revolutions originating inside static societies have been the result of ideas created by such a privileged person and not, as one might perhaps expect, by the people who were most oppressed.

Our own society did not become dynamic through the sudden disintegration of a static society, but through many generations of evolution. Uniquely, it has been characterised by a steady replacement of anti-rational memes by rational ones, a process which we may call the *Enlightenment*, the name usually given to the manifestations of the process in eighteenth-century philosophy and politics. When this process began is debatable, but it was certainly well before the eighteenth century. My own guess is that staticity was irreversibly destroyed by the discoveries of Galileo and Newton, which inaugurated modern science. In meme terms, these discoveries became rational memes of unprecedented resilience and fecundity, which the existing memes had no way of dislodging and whose replication has caused an avalanche of further changes. These include the extinction or near-extinction in our society of whole classes of anti-rational memes which have prospered in all other known societies, such as the memes which caused political tyranny, the subjugation of women and the torture of children.

Despite these successes, the Enlightenment is nowhere near complete. The scientific culture is dominated by rational memes; the conduct of politics also (though not the formation of people's political views); but in other areas of life, many memes are still replicated in the old, anti-rational manner, i.e. by means that violate the recipients' preferences and do not harness the recipients' creativity. Indeed, most of the constitutive memes of our society are close

variants of memes that first evolved under static-society conditions. Take marriage, for instance. The institution has been slightly modified by the toleration of divorce, homosexuality, extramarital sex etc. Nevertheless, the majority of people still enter into this obligation under the same, traditional terms. Although the complex behaviour that they have undertaken to enact will condition virtually every decision they subsequently make, very few marrying couples try to adapt the terms of the undertaking to their individual preferences. Most do not even consider what the consequences to themselves of enacting the required behaviour might be. Inevitably, many become unhappy with the consequences. Divorce is common – though typically, the holders of the marriage meme go on to enact it again, under the same terms. That people devote so much of their lives to the attempted enactment of complex, predefined behaviour which they do not feel able, or entitled, to adapt to their own purposes is testimony, not to "the triumph of hope over experience", as Dr Johnson joked, but to the continuing ability of an ancient meme to replicate itself. Less extreme examples of the same type of meme govern the employer-employee relationship. More extreme examples govern, as we should expect, many aspects of child rearing.

So our society is still very much in transition. The co-existence of rational and anti-rational memes makes this transition unstable because memes of each type cause behaviours that impede the replication of the other. Take, for example, the memes for religious faith and scientific rationality. People who are affected by both memes, but in whom religious faith predominates, try hard not to pass their doubts on to others. And people in whom scientific rationality predominates try hard not to pass on the vestiges of their belief in the supernatural. People in whom these two memes are of comparable strength cannot enact the behaviour dictated by either. Instead they enact a mixture of the behaviour that they cannot bring themselves to deviate from, and the behaviour that they rationally deem best. Such a mixture neither enforces literal accuracy nor cedes control to the receiver's preferences and creativity. Hence neither of the two possible ways of achieving stability operates. Unless the mixture of behaviours is itself a replicator (which would mean that the holder had invented a viable new religion – a very rare event), it must keep changing haphazardly from one generation to the next.

From the point of view of anti-rational memes, the current transitional era is one in which errors are already accumulating at a rate much greater than anti-rational evolution can cope with. The surviving anti-rational memes are being replicated by less effective variants of their traditional behaviours, so they are failing to preserve their implicit knowledge and are losing their functionality. As they become unable, one by one, to cause their own replication, they are becoming extinct. From the point of view of the rational memes, it is an era in which many entrenched obstacles to their evolution and survival are still dangerously active. The knowledge in rational memes is constantly under threat too. For to preserve its knowledge – i.e. itself – a rational meme must express itself in continually changing behaviours and concepts, and these changes require a flow of new ideas. If the supply and quality of such ideas is diminished by the operation of anti-rational memes, the knowledge in the rational meme may be lost. For example, the overthrow of a tyrant does not immediately destroy all copies of the tyrannical society's constitutive memes. They remain in people's minds and impede the functioning of any rational institutions that are set up. But those rational institutions, having initiated change, cause problems requiring further changes, not least in the institutions themselves. If they do not function well enough to solve those problems, the institutions will not survive. The old memes may reassert themselves, or worse variants may arise, or there may be chaos and the further destruction of knowledge.

However, few anti-rational memes are *wholly* harmful, like the peacock's tail. Most contain implicitly, among their harmful adaptations, some useful knowledge. Some of this knowledge is essential, not only to the static society in which it evolved, but also to the Enlightened society that we want to create. For example, moral principles have historically been transmitted anti-rationally, usually as part of a religious package. It is widely believed that nowadays, as religious belief is declining, people are behaving less morally and that is why, for instance, crime is increasing. Whether or not this is so, it serves as a simple illustration of something that must be happening all the time. Knowledge is being lost as anti-rational memes are superseded. When an anti-rational meme begins to lose its viability and we nevertheless need its knowledge, we are placed in the position of having to discover that knowledge for ourselves, either in the form of explicit theories or as new inexplicit ideas that are capable of being reliably transmitted. That is why we cannot complete the Enlightenment by a single revolutionary effort. We cannot, simply by desiring it, immediately create the knowledge that is required even to identify, far less to replace, all our existing anti-rational memes.

***Living with memes***

So anti-rational memes still form a substantial part of the personality of every member of our society. It follows that much of our behaviour and many of our conscious and unconscious thoughts are not chosen by us, nor are they good for us, individually or collectively. In biological language they are, at best, burdensome but partially symbiotic parasites. At worst they are deadly viruses with elaborate, painful symptoms that not only stultify their victims but force them to spend their lives infecting other people. Like slave-drivers inside our minds, they are often cruel, always indifferent to our wishes, yet sometimes – and this is both a bitter and a consoling fact – they provide us with the necessities of life.

I have used religions as my running example of anti-rational memes, but they are by no means the only examples in our society, nor the most important or harmful. I have chosen them because their replication strategies and life-cycles are uncontroversial and relatively simple: virtually all religions are openly anti-rational, commanding their members to suppress their critical faculties, enact the memes faithfully and propagate them accurately. But it is unusual for so much of an anti-rational meme to be explicit and conscious. As I have explained, we can expect anti-rational memes generally to hide their content from their holders. So it is difficult to discuss specific cases without running into controversy. But the purpose of this article is not to argue about individual memes. It is to discuss the theory of meme evolution in general. Although in any particular case it may be arguable that a given meme is rational, or that a given action on a given occasion is not being dictated by a meme at all, I do not see any escape from the case I have made that anti-rational memes exist in all of us: To lead our lives, each of us needs far more information than any one person can possibly create. Therefore it is inevitable that most of this information comes from other people, and consists of replicators – memes. These are subject to variation and selection, so they evolve. And in the circumstances under which many of them evolved, evolution favours those which, once installed, prevent themselves from being altered.

There are some similarities between these conclusions about anti-rational memes, and what some people believe about *genes.* Throughout modern times there has been controversy over the extent to which human behaviour is genetically influenced: the "nature versus nurture" controversy. It is indisputable that some of our behaviour – such as breathing – is caused by genes that we have inherited from our animal ancestors. The controversy is about the details of that inheritance, and in particular about two issues which have wide implications for the

human philosophies. The first is whether any of our genetic drives or predispositions are significant obstacles to rational thought and behaviour. The second is whether certain *differences* in human attributes, such as intelligence and sexual preferences, are caused by genetic variations.

If there were a genetically caused human behaviour that tended to disable people's creativity and critical faculties, as anti-rational memes do, it might be argued that the Enlightenment programme is doomed to fail. However, genetically caused behaviour is not necessarily immutable. On the contrary, even human behaviours in whose favour there is heavy biological selection pressure, and which are immutable in animals, and which might therefore be expected to be under the tightest genetic control, are in the event all fully reversible by memes, or by individual decisions. For example, it is commonplace for human beings, either for cultural reasons or for their own private reasons, to lose the desire to eat, or to mate, or to avoid pain (e.g. in sport or sexual masochism), or indeed to survive. If such fundamental biological drives produce, in humans, only behaviours and desires that can readily be unlearned when they conflict with memes, or with the transient objectives of individuals, then what genetically determined behaviours cannot be unlearned? Only those, such as our reflexes or digestion, over which our ideas have no physical control. Whatever is physically under the control of our ideas, we and our memes can and do vary for our own purposes. If that conflicts with the purposes of a gene, so much the worse for the gene.

It is ironic that genetic influences on behaviour, which are often assumed to be immutable and therefore responsible for the intractability of many human problems, turn out to be effectively insignificant, while certain *learned* behaviours, caused by anti-rational memes, though never entirely immutable, turn out to be the real cause of that intractability. But that is not surprising when we consider that anti-rational memes possess highly evolved adaptations with no other function than to cause immutable behaviour in humans, and to thwart their holders' best efforts to think themselves free, while our genes evolved only to control the behaviour of our beast-like ancestors. Behavioural genes have had virtually no experience of human thought as an adversary. Their evolution must have ceased as soon as our capacity for reason evolved, because any selection pressure that favoured behavioural genes would almost certainly also favour anti-rational memes for the same behaviour. These could evolve thousands of times as quickly, and would be much better able to cope with subsequent changes in conditions that required further evolution. As soon as there was a meme occupying any niche, the selection pressure favouring the analogous gene would disappear. There must also be continuous selection pressure on existing behavioural genes to become more flexible, or to be dropped altogether.

By the same argument, we cannot expect any of the differences between human cultures to depend on differently adapted behavioural genes. And if there are other genetically evolved differences in human behaviour – say, between men and women – they can be significant only as long as subsequent meme evolution happens to favour the maintenance of those differences. Genes for any behaviour that is physically under the control of human minds will be out-evolved by memes as soon as they conflict. There could still be random (i.e. un-evolved) genetically-caused differences in human minds, but it seems implausible that variations in attributes such as intelligence or sexual preference could fail to affect a gene or meme's chances of being replicated, and if they did, there would be selection pressure and the surviving variants would no longer be random.

There is a science of *socio-biology* (or evolutionary psychology) which attempts to explain behaviour "genetically" by mathematically analysing the evolution of hypothetical genes for

behavioural tendencies. Although such research has been successful at explaining complex patterns of behaviour in other species, I doubt it has produced anything useful regarding distinctly human behaviour. The current analysis shows why: Human behaviour is meme-based, not gene based. If proper account were taken of the evolutionary mechanisms of memes, socio-biology could predict some features of long-lived, static societies, and probably also of long-lived static cultures within dynamic societies. However, most socio-biological research to date has ignored memes altogether, or has assumed that they evolve like genes, and thus it has deprived itself of even this limited relevance to human affairs. To extend socio-biology to dynamic societies, one would have to predict the evolving content of rational memes, which, however, depends on the exercise of the holders' creativity. Predicting the outcome of a creative act is tantamount to exercising at least that amount of creativity oneself; and therefore to predict the behaviour of dynamic societies requires an investment of at least as much creativity as that entire society has exercised over the relevant period. That is why I doubt that socio-biology will make many useful predictions about people in dynamic societies.

That we are still, to a significant degree, the slaves of our memes, is an uncomfortable fact. Most of us would admit to having a hang-up or two, but in the main, we consider our behaviour to be determined by our own decisions, and our decisions by our own, consistent preferences. I suspect that this rational self-image is a recent development of our society, many of whose most powerful memes explicitly promote, and implicitly give effect to, values such as reason, progress and the pursuit of happiness. We naturally try to explain ourselves in terms of those values, and we are assisted by the unconscious and self-disguising nature of most anti-rational memes. In earlier societies, people would not have been seriously troubled by the proposition that most of their lives were spent enacting onerous rituals rather than pursuing their own goals. On the contrary, the degree to which a person's life was controlled by piety, duty, obedience to authority and so on, was the very measure by which people judged themselves and others. Children who asked why they were required to behave in one way rather than another, or to make sacrifices that did not seem to make sense, would be told "because I say so", and in due course they would give their children the same reply to the same question, never realising that they were giving the full explanation. (This is a curious type of anti-rational meme whose overt content is true, though its holders do not believe it!)

How ought we to react to the knowledge that our minds are not entirely our own? No differently, in principle, from how we react to the possibility of our being in error for any other reason. It should not come as a surprise to anyone who is committed to rationality in the conduct of human affairs, and in their own minds, that we can be badly mistaken in any of our ideas even when we feel most strongly that no alternative idea is conceivable. I am not calling for some sort of mental hygiene campaign to prevent anti-rational memes from being expressed or transmitted, as we would do with an infectious disease. For many of them contain knowledge that we throw away at our peril. Our objective ought not to be simply to abandon anti-rational memes, but to replace them, incrementally, by rational ideas. In any case, the Enlightenment is not so much a matter of public policy as of individual psychology. Here the history of the Enlightenment is perhaps misleading, for until quite recently it *has* been associated with great political reforms. For example, the arbitrary power of the monarch was progressively diminished and replaced by institutions that were open to new ideas and whose decision-making procedures depended on the consent of individuals. Laws regulating family life and sexual behaviour have also been reformed, with individual choice and consent as the central principle. But legislation was never what mainly constituted the Enlightenment in the sense in which I have defined it, namely, the replacement in people's

minds of anti-rational memes by rational ones. Had these psychological events not occurred first, the legislative reforms would not have been possible. On the other hand, once a viable rational meme had taken root, its spread was bound to force the political change sooner or later. So the main agents of the Enlightenment were not the reforming legislators, nor even the campaigners, but the creators of new, rational memes, and the pioneers who tried to enact them in preference to ancient, anti-rational memes, and to persuade others to do the same when this provoked anger, ridicule and worse.

Nowadays more than ever, the defeat of an anti-rational meme does not primarily, if at all, involve banning its enactment, or legally safeguarding what it used to suppress. It involves individuals deciding not to re-enact a traditional behaviour blindly, and starting to think for themselves. Breaking the circle by not putting themselves through what their parents and grandparents felt compelled to endure. Not inflicting on their children what was inflicted on them. And then persuading others, by argument and example, to do the same. If legislation is needed for the new ideas to be fully enacted, it follows, rather than precedes, the spread of the ideas themselves.

In biology, the modern theory of evolution not only explains the diverse adaptations of organisms, it also changes the ground rules about what counts as an explanation. In explaining a particular adaptation, it is no longer satisfactory to point to its function in the organism's physiology, or to its role in the survival of the species. Instead, the primary questions are: what is the replication strategy of the gene for this adaptation; and how did that gene evolve? Similarly in the human philosophies, traditional explanations of human behaviour have been in terms of people's intentions, or in terms of the function of that behaviour in a group such as a nation or class. If it is tempting to think of biological evolution as being driven by the "purposes" of species, which it could not possibly be, it is even more tempting to explain meme-driven behaviour as being caused by human beings pursuing their own purposes, a phenomenon which really does occur and really is significant. But now we see that when applied to culturally transmitted behaviour, such explanations are, at best, incomplete. For example, why do Public Schools force their inmates to study Latin? An answer in terms of individual intentions might be that the teachers believe that learning Latin improves their pupils' reasoning abilities. An answer in terms of social functionality might be that having had Latin lessons is a shared experience that adds cohesion to the upper echelons of society. However, whether or not Latin lessons have either of these functions, or any other function, this cannot logically amount to an *explanation*. A genuine explanation has to account for two things: how Latin came to be taught originally (which is straightforward in this case – it was once the standard international language), and how the idea of teaching it, when enacted by one generation, causes the next generation to enact it again. For however much Latin may benefit everyone, if it is taught in a way that makes the pupils less likely to have their children taught Latin, it will soon be taught no longer. And if it is taught in a way that makes the pupils unable to resist having their children taught Latin in the same way, then Latin will be taught indefinitely, *whatever good or harm that may do.*

Furthermore, the Latin-teaching meme must, like every other meme, adopt either a rational or an anti-rational strategy for "making the pupils unable to resist having their children taught Latin in the same way". That is, it must either fill them with enthusiasm for Latin, or create some psychological disorder that leads directly or indirectly, but inexorably, to distress at the thought of their own children not being taught Latin. Since the former strategy would presumably result their retaining an interest in Latin after leaving school, and since, in the event, that is seldom observed, we may tentatively infer that the meme uses the anti-rational strategy. (Actually my description of the meme's life-history is an over-simplification. It does

not replicate itself in a simple parent-child-parent sequence, as genes must. Rather, it induces an intricate pattern of interactions between children, parents, grandparents, teachers, school proprietors and the academic community. Nevertheless, the essence of that pattern must be as I have stated it.)

The misconception for which I have criticised socio-biology – i.e. that it either ignores memes or assumes that they evolve like genes – is also present in the prevailing theories and methods of most of the human philosophies, including psychology, ethics, political philosophy, sociology and anthropology. One cannot expect any of these subjects to make much progress until they begin to take the detailed effects of meme evolution into account. Consequently it seems to me that the underlying general theory of meme evolution needs urgently to be developed using, among other things, the same mathematical and empirical methods as are now used in biology.

But even setting all those academic disciplines on more fruitful paths may not be the most valuable application of the theory of meme evolution. I have argued that everyone in our society is struggling with some very harmful memes, and that society as a whole is in an unstable condition as a result. Because most anti-rational memes function largely at the unconscious level, merely understanding that they exist may give us a powerful weapon against them. And the more we know about them, the better for each of us individually and for society. We can no more directly perceive an anti-rational meme at work in our minds than we can directly see the optical blind spot in our field of vision. However, just as with the optical blind spot, there is nothing to prevent our using a combination of argument and observation to detect a meme indirectly. Whenever we find ourselves enacting a complex behaviour that has been accurately repeated from one generation (not necessarily in the biological sense) to the next, we should be suspicious. If we find that enacting this behaviour thwarts our efforts to attain our personal objectives, or is faithfully continued when the ostensible justifications for it disappear, we should become more suspicious. When we think we have detected an anti-rational meme, whether in ourselves or in others, the immediate task is to understand how it works. Only then can we create the knowledge both to defy it and to replace any beneficial functions it may have. Our future depends on whether enough of us have the will and creativity to liberate ourselves, in our own lives, from the tyranny of anti-rational memes, while rescuing the immense knowledge that they contain.

**Notes**

1. In Britain, "Public School" refers not to state schools but instead to the most prestigious private schools, including Eton and Harrow.

2. Be sure to read David Deutsch's life-altering books: *The Fabric of Reality* and *The Beginning of Infinity*

David Deutsch, 2023, 'The evolution of culture', written March 1994, first published 6th December 2023 on the Taking Children Seriously website at:
https://takingchildrenseriously.com/the-evolution-of-culture