2 THE REPLICATORS

In the beginning was simplicity. It is difficult enough explaining how even a simple universe began. I take it as agreed that it would be even harder to explain the sudden springing up, fully armed, of complex order — life, or a being capable of creating life. Darwin's theory of evolution by natural selection is satisfying because it shows us a way in which simplicity could change into complexity, how unordered atoms could group themselves into ever more complex patterns until they ended up manufacturing people. Darwin provides a solution, the only feasible one so far suggested, to the deep problem of our existence. I will try to explain the great theory in a more general way than is customary, beginning with the time before evolution itself began.

Darwin's ‘survival of the fittest’ is really a special case of a more general law *of survival of the stable*. The universe is populated by stable things. A stable thing is a collection of atoms that is permanent enough or common enough to deserve a name. It may be a unique collection of atoms, such as the Matterhorn, that lasts long enough to be worth naming. Or it may be a *class* of entities, such as rain drops, that come into existence at a sufficiently high rate to deserve a collective name, even if any one of them is short-lived. The things that we see around us, and which we think of as needing explanation — rocks, galaxies, ocean waves — are all, to a greater or lesser extent, stable patterns of atoms. Soap bubbles tend to be spherical because this is a stable configuration for thin films filled with gas. In a spacecraft, water is also stable in spherical globules, but on earth, where there is gravity, the stable surface for standing water is flat and horizontal. Salt crystals tend to be cubes because this is a stable way of packing sodium and chloride ions together. In the sun the simplest atoms of all, hydrogen atoms, are fusing to form helium atoms, because in the conditions that prevail there the helium configuration is more stable. Other even more complex atoms are being formed in {13} stars all over the universe, ever since soon after the ‘big bang’ which, according to the prevailing theory, initiated the universe. This is originally where the elements on our world came from.

Sometimes when atoms meet they link up together in chemical reaction to form molecules, which may be more or less stable. Such molecules can be very large. A crystal such as a diamond can be regarded as a single molecule, a proverbially stable one in this case, but also a very simple one since its internal atomic structure is endlessly repeated. In modern living organisms there are other large molecules which are highly complex, and their complexity shows itself on several levels. The haemoglobin of our blood is a typical protein molecule. It is built up from chains of smaller molecules, amino acids, each containing a few dozen atoms arranged in a precise pattern. In the haemoglobin molecule there are 574 amino acid molecules. These are arranged in four chains, which twist around each other to form a globular three-dimensional structure of bewildering complexity. A model of a haemoglobin molecule looks rather like a dense thorn bush. But unlike a real thorn bush it is not a haphazard approximate pattern but a definite invariant structure, identically repeated, with not a twig nor a twist out of place, over six thousand million million million times in an average human body. The precise thorn bush shape of a protein molecule such as haemoglobin is stable in the sense that two chains consisting of the same sequences of amino acids will tend, like two springs, to come to rest in exactly the same three-dimensional coiled pattern. Haemoglobin thorn bushes are springing into their ‘preferred’ shape in your body at a rate of about four hundred million million per second, and others are being destroyed at the same rate.

Haemoglobin is a modern molecule, used to illustrate the principle that atoms tend to fall into stable patterns. The point that is relevant here is that, before the coming of life on earth, some rudimentary evolution of molecules could have occurred by ordinary processes of physics and chemistry. There is no need to think of design or purpose or directedness. If a group of atoms in the presence of energy falls into a stable pattern it will tend to stay that way. The earliest form of natural selection was simply a selection of stable forms and a rejection of unstable ones. There is no mystery about this. It had to happen by definition.

From this, of course, it does not follow that you can explain the existence of entities as complex as man by exactly the same principles {14} on their own. It is no good taking the right number of atoms and shaking them together with some external energy till they happen to fall into the right pattern, and out drops Adam! You may make a molecule consisting of a few dozen atoms like that, but a man consists of over a thousand million million million million atoms. To try to make a man, you would have to work at your biochemical cocktail-shaker for a period so long that the entire age of the universe would seem like an eye-blink, and even then you would not succeed. This is where Darwin's theory, in its most general form, comes to the rescue. Darwin's theory takes over from where the story of the slow building up of molecules leaves off.

The account of the origin of life that I shall give is necessarily speculative; by definition, nobody was around to see what happened. There are a number of rival theories, but they all have certain features in common. The simplified account I shall give is probably not too far from the truth.(1)

We do not know what chemical raw materials were abundant on earth before the coming of life, but among the plausible possibilities are water, carbon dioxide, methane, and ammonia: all simple compounds known to be present on at least some of the other planets in our solar system. Chemists have tried to imitate the chemical conditions of the young earth. They have put these simple substances in a flask and supplied a source of energy such as ultraviolet light or electric sparks — artificial simulation of primordial lightning. After a few weeks of this, something interesting is usually found inside the flask: a weak brown soup containing a large number of molecules more complex than the ones originally put in. In particular, amino acids have been found — the building blocks of proteins, one of the two great classes of biological molecules. Before these experiments were done, naturally-occurring amino acids would have been thought of as diagnostic of the presence of life. If they had been detected on, say Mars, life on that planet would have seemed a near certainty. Now, however, their existence need imply only the presence of a few simple gases in the atmosphere and some volcanoes, sunlight, or thundery weather. More recently, laboratory simulations of the chemical conditions of earth before the coming of life have yielded organic substances called purines and pyrimidines. These are building blocks of the genetic molecule, DNA itself.

Processes analogous to these must have given rise to the ‘primeval soup’ which biologists and chemists believe constituted the seas {15} some three to four thousand million years ago. The organic substances became locally concentrated, perhaps in drying scum round the shores, or in tiny suspended droplets. Under the further influence of energy such as ultraviolet light from the sun, they combined into larger molecules. Nowadays large organic molecules would not last long enough to be noticed: they would be quickly absorbed and broken down by bacteria or other living creatures. But bacteria and the rest of us are latecomers, and in those days large organic molecules could drift unmolested through the thickening broth.

At some point a particularly remarkable molecule was formed by accident. We will call it the *Replicator*. It may not necessarily have been the biggest or the most complex molecule around, but it had the extraordinary property of being able to create copies of itself. This may seem a very unlikely sort of accident to happen. So it was. It was exceedingly improbable. In the lifetime of a man, things that are that improbable can be treated for practical purposes as impossible. That is why you will never win a big prize on the football pools. But in our human estimates of what is probable and what is not, we are not used to dealing in hundreds of millions of years. If you filled in pools coupons every week for a hundred million years you would very likely win several jackpots.

Actually a molecule that makes copies of itself is not as difficult to imagine as it seems at first, and it only had to arise once. Think of the replicator as a mould or template. Imagine it as a large molecule consisting of a complex chain of various sorts of building block molecules. The small building blocks were abundantly available in the soup surrounding the replicator. Now suppose that each building block has an affinity for its own kind. Then whenever a building block from out in the soup lands up next to a part of the replicator for which it has an affinity, it will tend to stick there. The building blocks that attach themselves in this way will automatically be arranged in a sequence that mimics that of the replicator itself. It is easy then to think of them joining up to form a stable chain just as in the formation of the original replicator. This process could continue as a progressive stacking up, layer upon layer. This is how crystals are formed.

On the other hand, the two chains might split apart, in which case we have two replicators, each of which can go on to make further copies.

A more complex possibility is that each building block has affinity not for its own kind, but reciprocally for one particular other kind. {16} Then the replicator would act as a template not for an identical copy, but for a kind of 'negative', which would in its turn remake an exact copy of the original positive. For our purposes it does not matter whether the original replication process was positive-negative or positive-positive, though it is worth remarking that the modern equivalents of the first replicator, the DNA molecules, use positive-negative replication. What does matter is that suddenly a new kind of 'stability' came into the world. Previously it is probable that no particular kind of complex molecule was very abundant in the soup, because each was dependent on building blocks happening to fall by luck into a particular stable configuration. As soon as the replicator was born it must have spread its copies rapidly throughout the seas, until the smaller building block molecules became a scarce resource, and other larger molecules were formed more and more rarely.

So we seem to arrive at a large population of identical replicas. But now we must mention an important property of any copying process: it is not perfect. Mistakes will happen. I hope there are no misprints in this book, but if you look carefully you may find one or two. They will probably not seriously distort the meaning of the sentences, because they will be 'first generation' errors. But imagine the days before printing, when books such as the Gospels were copied by hand. All scribes, however careful, are bound to make a few errors, and some are not above a little wilful 'improvement'. If they all copied from a single master original, meaning would not be greatly perverted. But let copies be made from other copies, which in their turn were made from other copies, and errors will start to become cumulative and serious. We tend to regard erratic copying as a bad thing, and in the case of human documents it is hard to think of examples where errors can be described as improvements. I suppose the scholars of the Septuagint could at least be said to have started something big when they mistranslated the Hebrew word for 'young woman' into the Greek word for 'virgin', coming up with the prophecy: 'Behold a virgin shall conceive and bear a son . . .'(2) Anyway, as we shall see, erratic copying in biological replicators can in a real sense give rise to improvement, and it was essential for the progressive evolution of life that some errors were made. We do not know how accurately the original replicator molecules made their copies. Their modern descendants, the DNA molecules, are astonishingly faithful compared with the most high-fidelity human {17} copying process, but even they occasionally make mistakes, and it is ultimately these mistakes that make evolution possible. Probably the original replicators were far more erratic, but in any case we may be sure that mistakes were made, and these mistakes were cumulative.

As mis-copyings were made and propagated, the primeval soup became filled by a population not of identical replicas, but of several varieties of replicating molecules, all ‘descended’ from the same ancestor. Would some varieties have been more numerous than others? Almost certainly yes. Some varieties would have been inherently more stable than others. Certain molecules, once formed, would be less likely than others to break up again. These types would become relatively numerous in the soup, not only as a direct logical consequence of their ‘longevity’, but also because they would have a long time available for making copies of themselves. Replicators of high longevity would therefore tend to become more numerous and, other things being equal, there would have been an ‘evolutionary trend’ towards greater longevity in the population of molecules.

But other things were probably not equal, and another property of a replicator variety that must have had even more importance in spreading it through the population was speed of replication or ‘fecundity’. If replicator molecules of type A make copies of themselves on average once a week while those of type B make copies of themselves once an hour, it is not difficult to see that pretty soon type A molecules are going to be far outnumbered, even if they ‘live’ much longer than B molecules. There would therefore probably have been an ‘evolutionary trend’ towards higher ‘fecundity’ of molecules in the soup. A third characteristic of replicator molecules which would have been positively selected is accuracy of replication. If molecules of type X and type Y last the same length of time and replicate at the same rate, but X makes a mistake on average every tenth replication while Y makes a mistake only every hundredth replication, Y will obviously become more numerous. The X contingent in the population loses not only the errant ‘children’ themselves, but also all their descendants, actual or potential.

If you already know something about evolution, you may find something slightly paradoxical about the last point. Can we reconcile the idea that copying errors are an essential prerequisite for evolution to occur, with the statement that natural selection favours high copying-fidelity? The answer is that although evolution may seem, in some vague sense, a ‘good thing’, especially since we are the product {18} of it, nothing actually ‘wants’ to evolve. Evolution is something that happens, willy-nilly, in spite of all the efforts of the replicators (and nowadays of the genes) to prevent it happening. Jacques Monod made this point very well in his Herbert Spencer lecture, after wryly remarking: ‘Another curious aspect of the theory of evolution is that everybody thinks he understands it!’

To return to the primeval soup, it must have become populated by stable varieties of molecule; stable in that either the individual molecules lasted a long time, or they replicated rapidly, or they replicated accurately. Evolutionary trends toward these three kinds of stability took place in the following sense: if you had sampled the soup at two different times, the later sample would have contained a higher proportion of varieties with high longevity/fecundity/copying-fidelity. This is essentially what a biologist means by evolution when he is speaking of living creatures, and the mechanism is the same — natural selection.

Should we then call the original replicator molecules ‘living’? Who cares? I might say to you ‘Darwin was the greatest man who has ever lived’, and you might say ‘No, Newton was’, but I hope we would not prolong the argument. The point is that no conclusion of substance would be affected whichever way our argument was resolved. The facts of the lives and achievements of Newton and Darwin remain totally unchanged whether we label them ‘great’ or not. Similarly, the story of the replicator molecules probably happened something like the way I am telling it, regardless of whether we choose to call them ‘living’. Human suffering has been caused because too many of us cannot grasp that words are only tools for our use, and that the mere presence in the dictionary of a word like ‘living’ does not mean it necessarily has to refer to something definite in the real world.

Whether we call the early replicators living or not, they were the ancestors of life; they were our founding fathers.

The next important link in the argument, one that Darwin himself laid stress on (although he was talking about animals and plants, not molecules) is *competition*. The primeval soup was not capable of supporting an infinite number of replicator molecules. For one thing, the earth's size is finite, but other limiting factors must also have been important. In our picture of the replicator acting as a template or mould, we supposed it to be bathed in a soup rich in the small building block molecules necessary to make copies. But when the replicators became numerous, building blocks must have been {19} used up at such a rate that they became a scarce and precious resource. Different varieties or strains of replicator must have competed for them. We have considered the factors that would have increased the numbers of favoured kinds of replicator. We can now see that less-favoured varieties must actually have become *less* numerous because of competition, and ultimately many of their lines must have gone extinct. There was a struggle for existence among replicator varieties. They did not know they were struggling, or worry about it; the struggle was conducted without any hard feelings, indeed without feelings of any kind. But they were struggling, in the sense that any mis-copying that resulted in a new higher level of stability, or a new way of reducing the stability of rivals, was automatically preserved and multiplied. The process of improvement was cumulative. Ways of increasing stability and of decreasing rivals' stability became more elaborate and more efficient. Some of them may even have 'discovered' how to break up molecules of rival varieties chemically, and to use the building blocks so released for making their own copies. These proto-carnivores simultaneously obtained food and removed competing rivals. Other replicators perhaps discovered how to protect themselves, either chemically, or by building a physical wall of protein around themselves. This may have been how the first living cells appeared. Replicators began not merely to exist, but to construct for themselves containers, vehicles for their continued existence. The replicators that survived were the ones that built *survival machines* for themselves to live in. The first survival machines probably consisted of nothing more than a protective coat. But making a living got steadily harder as new rivals arose with better and more effective survival machines. Survival machines got bigger and more elaborate, and the process was cumulative and progressive.

Was there to be any end to the gradual improvement in the techniques and artifices used by the replicators to ensure their own continuation in the world? There would be plenty of time for improvement. What weird engines of self-preservation would the millennia bring forth? Four thousand million years on, what was to be the fate of the ancient replicators? They did not die out, for they are past masters of the survival arts. But do not look for them floating loose in the sea; they gave up that cavalier freedom long ago. Now they swarm in huge colonies, safe inside gigantic lumbering robots,(3) sealed off from the outside world, communicating with it by tortuous {20} indirect routes, manipulating it by remote control. They are in you and in me; they created us, body and mind; and their preservation is the ultimate rationale for our existence. They have come a long way, those replicators. Now they go by the name of genes, and we are their survival machines.

6 GENESMANSHIP

What is the selfish gene? It is not just one single physical bit of DNA. Just as in the primeval soup, it is *all replicas* of a particular bit of DNA, distributed throughout the world. If we allow ourselves the licence of talking about genes as if they had conscious aims, always reassuring ourselves that we could translate our sloppy language back into respectable terms if we wanted to, we can ask the question, what is a single selfish gene trying to do? It is trying to get more numerous in the gene pool. Basically it does this by helping to program the bodies in which it finds itself to survive and to reproduce. But now we are emphasizing that 'it' is a distributed agency, existing in many different individuals at once. The key point of this chapter is that a gene might be able to assist *replicas* of itself that are sitting in other bodies. If so, this would appear as individual altruism but it would be brought about by gene selfishness.

Consider the gene for being an albino in man. In fact several genes exist that can give rise to albinism, but I am talking about just one of them. It is recessive; that is, it has to be present in double dose in order for the person to be an albino. This is true of about 1 in 20,000 of us. But it is also present, in single dose, in about 1 in 70 of us, and these individuals are not albinos. Since it is distributed in many individuals, a gene such as the albino gene could, in theory, assist its own survival in the gene pool by programming its bodies to behave altruistically towards other albino bodies, since these are known to contain the same gene. The albino gene should be quite happy if some of the bodies that it inhabits die, provided that in doing so they help other bodies containing the same gene to survive. If the albino gene could make one of its bodies save the lives often albino bodies, then even the death of the altruist is amply compensated by the increased numbers of albino genes in the gene pool.

Should we then expect albinos to be especially nice to each other? Actually the answer is probably no. In order to see why not, we must {89} temporarily abandon our metaphor of the gene as a conscious agent, because in this context it becomes positively misleading. We must translate back into respectable, if more longwinded terms. Albino genes do not really ‘want’ to survive or to help other albino genes. But if the albino gene just happened to cause its bodies to behave altruistically towards other albinos, then automatically, willy-nilly, it would tend to become more numerous in the gene pool as a result. But, in order for this to happen, the gene would have to have two independent effects on bodies. Not only must it confer its usual effect of a very pale complexion. It must also confer a tendency to be selectively altruistic towards individuals with a very pale complexion. Such a double-effect gene could, if it existed, be very successful in the population.

Now it is true that genes do have multiple effects, as I emphasized in Chapter 3. It is theoretically possible that a gene could arise which conferred an externally visible ‘label’, say a pale skin, or a green beard, or anything conspicuous, and also a tendency to be specially nice to bearers of that conspicuous label. It is possible, but not particularly likely. Green beardedness is just as likely to be linked to a tendency to develop ingrowing toenails or any other trait, and a fondness for green beards is just as likely to go together with an inability to smell freesias. It is not very probable that one and the same gene would produce both the right label and the right sort of altruism. Nevertheless, what may be called the Green Beard Altruism Effect is a theoretical possibility.

An arbitrary label like a green beard is just one way in which a gene might ‘recognize’ copies of itself in other individuals. Are there any other ways? A particularly direct possible way is the following. The possessor of an altruistic gene might be recognized simply by the fact that he does altruistic acts. A gene could prosper in the gene pool if it ‘said’ the equivalent of: ‘Body, *if A* is drowning as a result of trying to save someone else from drowning, jump in and rescue *A.’* The reason such a gene could do well is that there is a greater than average chance that *A* contains the same life-saving altruistic gene. The fact that *A* is seen to be trying to rescue somebody else is a label, equivalent to a green beard. It is less arbitrary than a green beard, but it still seems rather implausible. Are there any plausible ways in which genes might ‘recognize’ their copies in other individuals?

The answer is yes. It is easy to show that *close relatives* — kin — have a greater than average chance of sharing genes. It has long been clear that this must be why altruism by parents towards their young is so common. What R. A. Fisher, J. B. S. Haldane, and especially W. D. Hamilton realized, was that the same applies to other close relations — brothers and sisters, nephews and nieces, close cousins. If an individual dies in order to save ten close relatives, one copy of the kin-altruism gene may be lost, but a larger number of copies of the same gene is saved.

‘A larger number’ is a bit vague. So is ‘close relatives’. We can do better than that, as Hamilton showed. His two papers of 1964 are among the most important contributions to social ethology ever written, and I have never been able to understand why they have been so neglected by ethologists (his name does not even appear in the index of two major textbooks of ethology, both published in 1970).(1) Fortunately there are recent signs of a revival of interest in his ideas. Hamilton's papers are rather mathematical, but it is easy to grasp the basic principles intuitively, without rigorous mathematics, though at the cost of some over-simplification. The thing we want to calculate is the probability, or odds, that two individuals, say two sisters, share a particular gene.

For simplicity I shall assume that we are talking about genes that are rare in the gene pool as a whole.(2) Most people share ‘the gene for not being an albino’, whether they are related to each other or not. The reason this gene is so common is that in nature albinos are less likely to survive than non-albinos because, for example, the sun dazzles them and makes them relatively unlikely to see an approaching predator. We are not concerned with explaining the prevalence in the gene pool of such obviously ‘good’ genes as the gene for not being an albino. We are interested in explaining the success of genes specifically as a result of their altruism. We can therefore assume that, at least in the early stages of this process of evolution, these genes are rare. Now the important point is that even a gene that is rare in the population as a whole is common within a family. I contain a number of genes that are rare in the population as a whole, and you also contain genes that are rare in the population as a whole. The chance that we both contain the same rare genes is very small indeed. But the chances are good that my sister contains a particular rare gene that I contain, and the chances are equally good that your sister contains a rare gene in common with you. The odds are in this case exactly 50 per cent, and it is easy to explain why.

Suppose you contain one copy of the gene *G*. You must have {91} received it either from your father or from your mother (for convenience we can neglect various infrequent possibilities — that *G* is a new mutation, that both your parents had it, or that either of your parents had two copies of it). Suppose it was your father who gave you the gene. Then every one of his ordinary body cells contained one copy of *G*. Now you will remember that when a man makes a sperm he doles out half his genes to it. There is therefore a 50 per cent chance that the sperm that begot your sister received the gene *G*. If, on the other hand, you received *G* from your mother, exactly parallel reasoning shows that half of her eggs must have contained *G*; once again, the chances are 50 per cent that your sister contains *G*. This means that if you had 100 brothers and sisters, approximately 50 of them would contain any particular rare gene that you contain. It also means that if you have 100 rare genes, approximately 50 of them are in the body of any one of your brothers or sisters.

You can do the same kind of calculation for any degree of kinship you like. An important relationship is that between parent and child. If you have one copy of gene *H*, the chance that any particular one of your children has it is 50 per cent, because half your sex cells contain *H*, and any particular child was made from one of those sex cells. If you have one copy of gene *F*, the chance that your father also had *F* is 50 per cent, because you received half your genes from him, and half from your mother. For convenience we use an index of *relatedness*, which expresses the chance of a gene being shared between two relatives. The relatedness between two brothers is 1/2, since half the genes possessed by one brother will be found in the other. This is an average figure: by the luck of the meiotic draw, it is possible for particular pairs of brothers to share more or fewer genes than this. The relatedness between parent and child is always exactly 1/2

It is rather tedious going through the calculations from first principles every time, so here is a rough and ready rule for working out the relatedness between any two individuals *A* and *B*. You may find it useful in making your will, or in interpreting apparent resemblances in your own family. It works for all simple cases, but breaks down where incestuous mating occurs, and in certain insects, as we shall see.

First identify all the *common ancestors of A* and *B*. For instance, the common ancestors of a pair of first cousins are their shared grandfather and grandmother. Once you have found a common ancestor, it is of course logically true that all his ancestors are common to *A* and {92} *B* as well. However, we ignore all but the most recent common ancestors. In this sense, first cousins have only two common ancestors. If *B* is a lineal descendant of *A*, for instance his great grandson, then *A* himself is the 'common ancestor' we are looking for.

Having located the common ancestor(s) of *A* and *B*, count the *generation distance* as follows. Starting at *A*, climb up the family tree until you hit a common ancestor, and then climb down again to *B*. The total number of steps up the tree and then down again is the generation distance. For instance, if *A* is *B*'s uncle, the generation distance is 3. The common ancestor is *A*'s father (say) and *B*'s grandfather. Starting at *A* you have to climb up one generation in order to hit the common ancestor. Then to get down to *B* you have to descend two generations on the other side. Therefore the generation distance is $1 + 2 = 3$.

Having found the generation distance between *A* and *B* via a particular common ancestor, calculate that part of their relatedness for which that ancestor is responsible. To do this, multiply 1/2 by itself once for each step of the generation distance. If the generation distance is 3, this means calculate $1/2 \times 1/2 \times 1/2$ or $(1/2)^3$. If the generation distance via a particular ancestor is equal to g steps, the portion of relatedness due to that ancestor is $(1/2)^g$.

But this is only part of the relatedness between *A* and *B*. If they have more than one common ancestor we have to add on the equivalent figure for each ancestor. It is usually the case that the generation distance is the same for all common ancestors of a pair of individuals. Therefore, having worked out the relatedness between *A* and *B* due to any one of the ancestors, all you have to do in practice is to multiply by the number of ancestors. First cousins, for instance, have two common ancestors, and the generation distance via each one is 4. Therefore their relatedness is $2 \times (1/2)^4 = 1/8$. *If A* is *B*'s great-grandchild, the generation distance is 3 and the number of common 'ancestors' is 1 (*B* himself), so the relatedness is $1 \times (1/2)^3 = 1/8$. Genetically speaking, your first cousin is equivalent to a great-grandchild. Similarly, you are just as likely to 'take after' your uncle (relatedness = $2 \times (1/2)^3 = 1/4$) as after your grandfather (relatedness = $1 \times (1/2)^2 = 1/4$).

For relationships as distant as third cousin ($2 \times (1/2)^8 = 1/128$), we are getting down near the baseline probability that a particular gene possessed by *A* will be shared by any random individual taken from the population. A third cousin is not far from being equivalent to any {93} old Tom, Dick, or Harry as far as an altruistic gene is concerned. A second cousin (relatedness = 1/32) is only a little bit special; a first cousin somewhat more so (1/8). Full brothers and sisters, and parents and children are very special (1/2), and identical twins (relatedness = 1) just as special as oneself. Uncles and aunts, nephews and nieces, grandparents and grandchildren, and half brothers and half sisters, are intermediate with a relatedness of 1/4.

Now we are in a position to talk about genes for kin-altruism much more precisely. A gene for suicidally saving five cousins would not become more numerous in the population, but a gene for saving five brothers or ten first cousins would. The minimum requirement for a suicidal altruistic gene to be successful is that it should save more than two siblings (or children or parents), or more than four half-siblings (or uncles, aunts, nephews, nieces, grandparents, grand-children), or more than eight first cousins, etc. Such a gene, on average, tends to live on in the bodies of enough individuals saved by the altruist to compensate for the death of the altruist itself.

If an individual could be sure that a particular person was his identical twin, he should be exactly as concerned for his twin's welfare as for his own. Any gene for twin altruism is bound to be carried by both twins, therefore if one dies heroically to save the other the gene lives on. Nine-banded armadillos are born in a litter of identical quadruplets. As far as I know, no feats of heroic self-sacrifice have been reported for young armadillos, but it has been pointed out that some strong altruism is definitely to be expected, and it would be well worth somebody's while going out to South America to have a look.(3)

We can now see that parental care is just a special case of kin altruism. Genetically speaking, an adult should devote just as much care and attention to its orphaned baby brother as it does to one of its own children. Its relatedness to both infants is exactly the same, 1/2. In gene selection terms, a gene for big sister altruistic behaviour should have just as good a chance of spreading through the population as a gene for parental altruism. In practice, this is an over-simplification for various reasons which we shall come to later, and brotherly or sisterly care is nothing like so common in nature as parental care. But the point I am making here is that there is nothing special *genetically* speaking about the parent/child relationship as against the brother/sister relationship. The fact that parents actually hand on genes to children, but sisters do not hand on genes to each other is irrelevant, {94} since the sisters both receive identical replicas of the same genes from the same parents.

Some people use the term *kin selection* to distinguish this kind of natural selection from group selection (the differential survival of groups) and individual selection (the differential survival of individuals). Kin selection accounts for within-family altruism; the closer the relationship, the stronger the selection. There is nothing wrong with this term, but unfortunately it may have to be abandoned because of recent gross misuses of it, which are likely to muddle and confuse biologists for years to come. E. O. Wilson, in his otherwise admirable *Sociobiology: The New Synthesis*, defines kin selection as a special case of group selection. He has a diagram which clearly shows that he thinks of it as intermediate between ‘individual selection’, and ‘group selection’ in the conventional sense — the sense that I used in Chapter 1. Now group selection — even by Wilson's own definition — means the differential survival of *groups* of individuals. There is, to be sure, a sense in which a family is a special kind of group. But the whole point of Hamilton's argument is that the distinction between family and non-family is not hard and fast, but a matter of mathematical probability. It is no part of Hamilton's theory that animals should behave altruistically towards all ‘members of the family’, and selfishly to everybody else. There are no definite lines to be drawn between family and non-family. We do not have to decide whether, say, second cousins should count as inside the family group or outside it: we simply expect that second cousins should be 1/16 as likely to receive altruism as offspring or siblings.

9 BATTLE OF THE SEXES

If there is conflict of interest between parents and children, who share 50 per cent of each others' genes, how much more severe must be the conflict between mates, who are not related to each other?(1) All that they have in common is a 50 per cent genetic shareholding in the same children. Since father and mother are both interested in the welfare of different halves of the same children, there may be some advantage for both of them in cooperating with each other in rearing those children. If one parent can get away with investing less than his or her fair share of costly resources in each child, however, he will be better off, since he will have more to spend on other children by other sexual partners, and so propagate more of his genes. Each partner can therefore be thought of as trying to exploit the other, trying to force the other one to invest more. Ideally, what an individual would 'like' (I don't mean physically enjoy, although he might) would be to copulate with as many members of the opposite sex as possible, leaving the partner in each case to bring up the children. As we shall see, this state of affairs is achieved by the males of a number of species, but in other species the males are obliged *to* share an equal part of the burden of bringing up children. This view of sexual partnership, as a relationship of mutual mistrust and mutual exploitation, has been stressed especially by Trivers. It is a comparatively new one to ethologists. We had usually thought of sexual behaviour, copulation, and the courtship that precedes it, as essentially a cooperative venture undertaken for mutual benefit, or even for the good of the species!

Let us go right back to first principles, and inquire into the fundamental nature of maleness and femaleness. In Chapter 3 we discussed sexuality without stressing its basic asymmetry. We simply accepted that some animals are called male, and others female, without asking what these words really meant But what is the essence of maleness? What, at bottom, defines a female? We as {141} mammals see the sexes defined by whole syndromes of characteristics — possession of a penis, bearing of the young, suckling by means of special milk glands, certain chromosomal features, and so on. These criteria for judging the sex of an individual are all very well for mammals but, for animals and plants generally, they are no more reliable than is the tendency to wear trousers as a criterion for judging human sex. In frogs, for instance, neither sex has a penis. Perhaps, then, the words male and female have no general meaning. They are, after all, only words, and if we do not find them helpful for describing frogs, we are quite at liberty to abandon them. We could arbitrarily divide frogs into Sex 1 and Sex 2 if we wished. However, there is one fundamental feature of the sexes which can be used to label males as males, and females as females, throughout animals and plants. This is that the sex cells or 'gametes' of males are much smaller and more numerous than the gametes of females. This is true whether we are dealing with animals or plants. One group of individuals has large sex cells, and it is convenient to use the word female for them. The other group, which it is convenient to call male, has small sex cells. The difference is especially pronounced in reptiles and in birds, where a single egg cell is big enough and nutritious enough to feed a developing baby for several weeks. Even in humans, where the egg is microscopic, it is still many times larger than the sperm. As we shall see, it is possible to interpret all the other differences between the sexes as stemming from this one basic difference.

In certain primitive organisms, for instance some fungi, maleness and femaleness do not occur, although sexual reproduction of a kind does. In the system known as isogamy the individuals are not distinguishable into two sexes. Anybody can mate with anybody else. There are not two different sorts of gametes — sperms and eggs — but all sex cells are the same, called isogametes. New individuals are formed by the fusion of two isogametes, each produced by meiotic division. If we have three isogametes, *A*, *B*, and *C, A* could fuse with *B* or *C, B* could fuse with *A* or *C*. The same is never true of normal sexual systems. If *A* is a sperm and it can fuse with *B* or *C*, then *B* and *C* must be eggs and *B* cannot fuse with *C*.

When two isogametes fuse, both contribute equal numbers of genes to the new individual, and they also contribute equal amounts of food reserves. Sperms and eggs too contribute equal numbers of genes, but eggs contribute far more in the way of food reserves: {142} indeed, sperms make no contribution at all and are simply concerned with transporting their genes as fast as possible to an egg. At the moment of conception, therefore, the father has invested less than his fair share (i.e. 50 per cent) of resources in the offspring. Since each sperm is so tiny, a male can afford to make many millions of them every day. This means he is potentially able to beget a very large number of children in a very short period of time, using different females. This is only possible because each new embryo is endowed with adequate food by the mother in each case. This therefore places a limit on the number of children a female can have, but the number of children a male can have is virtually unlimited. Female exploitation begins here.(2)

Parker and others showed how this asymmetry might have evolved from an originally isogamous state of affairs. In the days when all sex cells were interchangeable and of roughly the same size, there would have been some that just happened to be slightly bigger than others. In some respects a big isogamete would have an advantage over an average-sized one, because it would get its embryo off to a good start by giving it a large initial food supply. There might therefore have been an evolutionary trend towards larger gametes. But there was a catch. The evolution of isogametes that were larger than was strictly necessary would have opened the door to selfish exploitation. Individuals who produced *smaller* than average gametes could cash in, provided they could ensure that their small gametes fused with extra-big ones. This could be achieved by making the small ones more mobile, and able to seek out large ones actively. The advantage to an individual of producing small, rapidly moving gametes would be that he could afford to make a larger number of gametes, and therefore could potentially have more children. Natural selection favoured the production of sex cells that were small and that actively sought out big ones to fuse with. So we can think of two divergent sexual 'strategies' evolving. There was the large-investment or 'honest' strategy. This automatically opened the way for a small-investment exploitative strategy. Once the divergence between the two strategies had started, it would have continued in runaway fashion. Medium-sized intermediates would have been penalized, because they did not enjoy the advantages of either of the two more extreme strategies. The exploiters would have evolved smaller and smaller size, and faster mobility. The honest ones would have evolved larger and larger size, to compensate for the ever-smaller {143} investment contributed by the exploiters, and they became immobile because they would always be actively chased by the exploiters anyway. Each honest one would 'prefer' to fuse with another honest one. But the selection pressure to lock out exploiters would have been weaker than the pressure on exploiters to duck under the barrier: the exploiters had more to lose, and they therefore won the evolutionary battle. The honest ones became eggs, and the exploiters became sperms.

Males, then, seem to be pretty worthless fellows, and on simple ‘good of the species’ grounds, we might expect that males would become less numerous than females. Since one male can theoretically produce enough sperms to service a harem of 100 females we might suppose that females should outnumber males in animal populations by 100 to 1. Other ways of putting this are that the male is more ‘expendable’, and the female more ‘valuable’ to the species. Of course, looked at from the point of view of the species as a whole, this is perfectly true. To take an extreme example, in one study of elephant seals, 4 per cent of the males accounted for 88 per cent of all the copulations observed. In this case, and in many others, there is a large surplus of bachelor males who probably never get a chance to copulate in their whole lives. But these extra males live otherwise normal lives, and they eat up the population's food resources no less hungrily than other adults. From a ‘good of the species’ point of view this is horribly wasteful; the extra males might be regarded as social parasites. This is just one more example of the difficulties that the group selection theory gets into. The selfish gene theory, on the other hand, has no trouble in explaining the fact that the numbers of males and females tend to be equal, even when the males who actually reproduce may be a small fraction of the total number. The explanation was first offered by R. A. Fisher.

The problem of how many males and how many females are born is a special case of a problem in parental strategy. Just as we discussed the optimal family size for an individual parent trying to maximize her gene survival, we can also discuss the optimal sex ratio. Is it better to entrust your precious genes to sons or to daughters? Suppose a mother invested all her resources in sons, and therefore had none left to invest in daughters: would she on average contribute more to the gene pool of the future than a rival mother who invested in daughters? Do genes for preferring sons become more or less numerous than genes for preferring daughters? What Fisher showed {144} is that under normal circumstances the stable sex ratio is 50:50. In order to see why, we must first know a little bit about the mechanics of sex determination.

In mammals, sex is determined genetically as follows. All eggs are capable of developing into either a male or a female. It is the sperms that carry the sex-determining chromosomes. Half the sperms produced by a man are female-producing, or X-sperms, and half are male-producing, or Y-sperms. The two sorts of sperms look alike. They differ with respect to one chromosome only. A gene for making a father have nothing but daughters could achieve its object by making him manufacture nothing but X-sperms. A gene for making a mother have nothing but daughters could work by making her secrete a selective spermicide, or by making her abort male embryos. What we seek is something equivalent to an evolutionarily stable strategy (ESS), although here, even more than in the chapter on aggression, strategy is just a figure of speech. An individual cannot literally choose the sex of his children. But genes for tending to have children of one sex or the other are possible. If we suppose that such genes, favouring unequal sex ratios, exist, are any of them likely to become more numerous in the gene pool than their rival alleles, which favour an equal sex ratio?

Suppose that in the elephant seals mentioned above, a mutant gene arose that tended to make parents have mostly daughters. Since there is no shortage of males in the population, the daughters would have no trouble finding mates, and the daughter-manufacturing gene could spread. The sex ratio in the population might then start to shift towards a surplus of females. From the point of view of the good of the species, this would be all right, because just a few males are quite capable of providing all the sperms needed for even a huge surplus of females, as we have seen. Superficially, therefore, we might expect the daughter-producing gene to go on spreading until the sex ratio was so unbalanced that the few remaining males, working flat out, could just manage. But now, think what an enormous genetic advantage is enjoyed by those few parents who have sons. Anyone who invests in a son has a very good chance of being the grandparent of hundreds of seals. Those who are producing nothing but daughters are assured of a safe few grandchildren, but this is nothing compared to the glorious genetic possibilities that open up before anyone specializing in sons. Therefore genes for {145} producing sons will tend to become more numerous, and the pendulum will swing back.

For simplicity I have talked in terms of a pendulum swing. In practice the pendulum would never have been allowed to swing that far in the direction of female domination, because the pressure to have sons would have started to push it back as soon as the sex ratio became unequal. The strategy of producing equal numbers of sons and daughters is an evolutionarily stable strategy, in the sense that any gene for departing from it makes a net loss.

I have told the story in terms of numbers of sons versus numbers of daughters. This is to make it simple, but strictly it should be worked out in terms of parental investment, meaning all the food and other resources that a parent has to offer, measured in the way discussed in the previous chapter. Parents should *invest* equally in sons and daughters. This usually means they should have numerically as many sons as they have daughters. But there could be unequal sex ratios that were evolutionarily stable, provided correspondingly unequal amounts of resources were invested in sons and daughters. In the case of the elephant seals, a policy of having three times as many daughters as sons, but of making each son a supermale by investing three times as much food and other resources in him, could be stable. By investing more food in a son and making him big and strong, a parent might increase his chances of winning the supreme prize of a harem. But this is a special case. Normally the amount invested in each son will roughly equal the amount invested in each daughter, and the sex ratio, in terms of numbers, is usually one to one.

In its long journey down the generations therefore, an average gene will spend approximately half its time sitting in male bodies, and the other half sitting in female bodies. Some gene effects show themselves only in bodies of one sex. These are called sex-limited gene effects. A gene controlling penis-length expresses this effect only in male bodies, but it is carried about in female bodies too and may have some quite different effect on female bodies. There is no reason why a man should not inherit a tendency to develop a long penis from his mother.

In whichever of the two sorts of body it finds itself, we can expect a gene to make the best use of the opportunities offered by that sort of body. These opportunities may well differ according to whether the body is male or female. As a convenient approximation, we can once again assume that each individual body is a selfish machine, trying to {146} do the best for all its genes. The best policy for such a selfish machine will often be one thing if it is male, and quite a different thing if it is female. For brevity, we shall again use the convention of thinking of the individual as though it had a conscious purpose. As before, we shall hold in the back of our mind that this is just a figure of speech. A body is really a machine blindly programmed by its selfish genes.

Consider again the mated pair with which we began the chapter. Both partners, as selfish machines, 'want' sons and daughters in equal numbers. To this extent they agree. Where they disagree is in who is going to bear the brunt of the cost of rearing each one of those children. Each individual wants as many surviving children as possible. The less he or she is obliged to invest in any one of those children, the more children he or she can have. The obvious way to achieve this desirable state of affairs is to induce your sexual partner to invest more than his or her fair share of resources in each child, leaving you free to have other children with other partners. This would be a desirable strategy for either sex, but it is more difficult for the female to achieve. Since she starts by investing more than the male, in the form of her large, food-rich egg, a mother is already at the moment of conception 'committed' to each child more deeply than the father is. She stands to lose more if the child dies than the father does. More to the point, she would have to invest more than the father *in the future* in order to bring a new substitute child up to the same level of development. If she tried the tactic of leaving the father holding the baby, while she went off with another male, the father might, at relatively small cost to himself, retaliate by abandoning the baby too. Therefore, at least in the early stages of child development, if any abandoning is going to be done, it is likely to be the father who abandons the mother rather than the other way around. Similarly, females can be expected to invest more in children than males, not only at the outset, but throughout development. So, in mammals for example, it is the female who incubates the foetus in her own body, the female who makes the milk to suckle it when it is born, the female who bears the brunt of the load of bringing it up and protecting it. The female sex is exploited, and the fundamental evolutionary basis for the exploitation is the fact that eggs are larger than sperms.

Of course in many species the father does work hard and faithfully at looking after the young. But even so, we must expect that there will {147} normally be some evolutionary pressure on males to invest a little bit less in each child, and to try to have more children by different wives. By this I simply mean that there will be a tendency for genes that say ‘Body, if you are male leave your mate a little bit earlier than my rival allele would have you do, and look for another female’, to be successful in the gene pool. The extent to which this evolutionary pressure actually prevails in practice varies greatly from species to species. In many, for example in the birds of paradise, the female receives no help at all from any male, and she rears her children on her own. Other species such as kittiwakes form monogamous pair-bonds of exemplary fidelity, and both partners cooperate in the work of bringing up children. Here we must suppose that some evolutionary counter-pressure has been at work: there must be a penalty attached to the selfish mate-exploitation strategy as well as a benefit, and in kittiwakes the penalty outweighs the benefit. It will in any case only pay a father to desert his wife and child if the wife has a reasonable chance of rearing the child on her own.

Trivers has considered the possible courses of action open to a mother who has been deserted by her mate. Best of all for her would be to try to deceive another male into adopting her child, ‘thinking’ it is his own. This might not be too difficult if it is still a foetus, not yet born. Of course, while the child bears half her genes, it bears no genes at all from the gullible step-father. Natural selection would severely penalize such gullibility in males and indeed would favour males who took active steps to kill any potential step-children as soon as they mated with a new wife. This is very probably the explanation of the so-called Bruce effect: male mice secrete a chemical which when smelt by a pregnant female can cause her to abort. She only aborts if the smell is different from that of her former mate. In this way a male mouse destroys his potential step-children, and renders his new wife receptive to his own sexual advances. Ardrey, incidentally, sees the Bruce effect as a population control mechanism! A similar example is that of male lions, who, when newly arrived in a pride, sometimes murder existing cubs, presumably because these are not their own children.

A male can achieve the same result without necessarily killing step-children. He can enforce a period of prolonged courtship before he copulates with a female, driving away all other males who approach her, and preventing her from escaping. In this way he can wait and see whether she is harbouring any little step-children in her {148} womb, and desert her if so. We shall see below a reason why a female might want a long ‘engagement’ period before copulation. Here we have a reason why a male might want one too. Provided he can isolate her from all contact with other males, it helps to avoid being the unwitting benefactor of another male's children.

Assuming then that a deserted female cannot fool a new male into adopting her child, what else can she do? Much may depend on how old the child is. If it is only just conceived, it is true that she has invested the whole of one egg in it and perhaps more, but it may still pay her to abort it and find a new mate as quickly as possible. In these circumstances it would be to the mutual advantage both of her and of the potential new husband that she should abort — since we are assuming she has no hope of fooling him into adopting the child. This could explain why the Bruce effect works from the female's point of view.

Another option open to a deserted female is to stick it out, and try and rear the child on her own. This will especially pay her if the child is already quite old. The older he is the more has already been invested in him, and the less it will take out of her to finish the job of rearing him. Even if he is still quite young, it might yet pay her to try to salvage something from her initial investment, even if she has to work twice as hard to feed the child, now that the male has gone. It is no comfort to her that the child contains half the male's genes too, and that she could spite him by abandoning it. There is no point in spite for its own sake. The child carries half her genes, and the dilemma is now hers alone.

Paradoxically, a reasonable policy for a female who is in danger of being deserted might be to walk out on the male *before* he walks out on her. This could pay her, even if she has already invested more in the child than the male has. The unpleasant truth is that in some circumstances an advantage accrues to the partner who deserts first, whether it is the father or the mother. As Trivers puts it, the partner who is left behind is placed in a cruel bind. It is a rather horrible but very subtle argument. A parent may be expected to desert, the moment it is possible for him or her to say the following: ‘This child is now far enough developed that either of us *could* finish off rearing it on our own. Therefore it would pay me to desert now, provided I could be sure my partner would not desert as well. If I did desert now, my partner would do whatever is best for her/his genes. He/she would be forced into making a more drastic decision than I am {149} making now, because I would have already left. My partner would "know" that if he/she left as well, the child would surely die. Therefore, assuming that my partner will take the decision that is best for his/her own selfish genes, I conclude that my own best course of action is to desert first. This is especially so, since my partner may be "thinking" along exactly the same lines, and may seize the initiative at any minute by deserting me!’ As always, the subjective soliloquy is intended for illustration only. The point is that genes for deserting first could be favourably selected simply because genes for deserting *second* would not be.

We have looked at some of the things that a female might do if she has been deserted by her mate. But these all have the air of making the best of a bad job. Is there anything a female can do to reduce the extent to which her mate exploits her in the first place? She has a strong card in her hand. She can refuse to copulate. She is in demand, in a seller's market. This is because she brings the dowry of a large, nutritious egg. A male who successfully copulates gains a valuable food reserve for his offspring. The female is potentially in a position to drive a hard bargain before she copulates. Once she has copulated she has played her ace — her egg has been committed to the male. It is all very well to talk about driving hard bargains, but we know very well it is not really like that. Is there any realistic way in which something equivalent to driving a hard bargain could evolve by natural selection? I shall consider two main possibilities, called the domestic-bliss strategy, and the he-man strategy.

The simplest version of the domestic-bliss strategy is this. The female looks the males over, and tries to spot signs of fidelity and domesticity in advance. There is bound to be variation in the population of males in their predisposition to be faithful husbands. If females could recognize such qualities in advance, they could benefit themselves by choosing males possessing them. One way for a female to do this is to play hard to get for a long time, to be coy. Any male who is not patient enough to wait until the female eventually consents to copulate is not likely to be a good bet as a faithful husband. By insisting on a long engagement period, a female weeds out casual suitors, and only finally copulates with a male who has proved his qualities of fidelity and perseverance in advance. Feminine coyness is in fact very common among animals, and so are prolonged courtship or engagement periods. As we have already seen, a long engagement can also benefit a male where {150} there is a danger of his being duped into caring for another male's child.

Courtship rituals often include considerable pre-copulation investment by the male. The female may refuse to copulate until the male has built her a nest. Or the male may have to feed her quite substantial amounts of food. This, of course, is very good from the female's point of view, but it also suggests another possible version of the domestic-bliss strategy. Could females force males to invest so heavily in their offspring *before* they allow copulation that it would no longer pay the males to desert *after* copulation? The idea is appealing. A male who waits for a coy female eventually to copulate with him is paying a cost: he is forgoing the chance to copulate with other females, and he is spending a lot of time and energy in courting her. By the time he is finally allowed to copulate with a particular female, he will inevitably be heavily ‘committed’ to her. There will be little temptation for him to desert her, if he knows that any future female he approaches will also procrastinate in the same manner before she will get down to business.

As I showed in a paper, there is a mistake in Trivers's reasoning here. He thought that prior investment in itself committed an individual to future investment. This is fallacious economics. A business man should never say ‘I have already invested so much in the Concorde airliner (for instance) that I cannot afford to scrap it now.’ He should always ask instead whether it would pay him *in the future*, to cut his losses, and abandon the project now, even though he has already invested heavily in it. Similarly, it is no use a female forcing a male to invest heavily in her in the hope that this, on its own, will deter the male from subsequently deserting. This version of the domestic-bliss strategy depends upon one further crucial assumption. This is that a majority of the females can be relied upon to play the same game. If there are loose females in the population, prepared to welcome males who have deserted their wives, then it could pay a male to desert his wife, no matter how much he has already invested in her children.

Much therefore depends on how the majority of females behave. If we were allowed to think in terms of a conspiracy of females there would be no problem. But a conspiracy of females can no more evolve than the conspiracy of doves which we considered in Chapter 5. Instead, we must look for evolutionarily stable strategies. Let us take Maynard Smith's method of analysing aggressive contests, {151} and apply it to sex.(3) It will be a little bit more complicated than the case of the hawks and doves, because we shall have two female strategies and two male strategies.

As in Maynard Smith's studies, the word ‘strategy’ refers to a blind unconscious behaviour program. Our two female strategies will be called *coy* and *fast*, and the two male strategies will be called *faithful and philanderer*. The behavioural rules of the four types are as follows. Coy females will not copulate with a male until he has gone through a long and expensive courtship period lasting several weeks. Fast females will copulate immediately with anybody. Faithful males are prepared to go on courting for a long time, and after copulation they stay with the female and help her to rear the young. Philanderer males lose patience quickly if a female will not copulate with them straight away: they go off and look for another female; after copulation too they do not stay and act as good fathers, but go off in search of fresh females. As in the case of the hawks and doves, these are not the only possible strategies, but it is illuminating to study their fates nevertheless.

Like Maynard Smith, we shall use some arbitrary hypothetical values for the various costs and benefits. To be more general it can be done with algebraic symbols, but numbers are easier to understand. Suppose that the genetic pay-off gained by each parent when a child is reared successfully is +15 units. The cost of rearing one child, the cost of all its food, all the time spent looking after it, and all the risks taken on its behalf, is –20 units. The cost is expressed as negative, because it is ‘paid out’ by the parents. Also negative is the cost of wasting time in prolonged courtship. Let this cost be –3 units.

Imagine we have a population in which all the females are coy, and all the males are faithful. It is an ideal monogamous society. In each couple, the male and the female both get the same average pay-off. They get +15 for each child reared; they share the cost of rearing it (–20) equally between the two of them, an average of –10 each. They both pay the –3 point penalty for wasting time in prolonged courtship. The average pay-off for each is therefore +15 – 10 – 3 = +2.

Now suppose a single fast female enters the population. She does very well. She does not pay the cost of delay, because she does not indulge in prolonged courtship. Since all the males in the population are faithful, she can reckon on finding a good father for her children whoever she mates with. Her average pay-off per child is {152} +15 – 10 = +5. She is 3 units better off than her coy rivals. Therefore fast genes will start to spread.

If the success of fast females is so great that they come to predominate in the population, things will start to change in the male camp too. So far, faithful males have had a monopoly. But now if a philanderer male arises in the population, he starts to do better than his faithful rivals. In a population where all the females are fast, the pickings for a philanderer male are rich indeed. He gets the +15 points if a child is successfully reared, and he pays neither of the two costs. What this lack of cost mainly means to him is that he is free to go off and mate with new females. Each of his unfortunate wives struggles on alone with the child, paying the entire –20 point cost, although she does not pay anything for wasting time in courting. The net pay-off for a fast female when she encounters a philanderer male is +15 – 20 = –5; the pay-off to the philanderer himself is +15. In a population in which all the females are fast, philanderer genes will spread like wildfire.

If the philanderers increase so successfully that they come to dominate the male part of the population, the fast females will be in dire straits. Any coy female would have a strong advantage. If a coy female encounters a philanderer male, no business results. She insists on prolonged courtship; he refuses and goes off in search of another female. Neither partner pays the cost of wasting time. Neither gains anything either, since no child is produced. This gives a net pay-off of zero for a coy female in a population where all the males are philanderers. Zero may not seem much, but it is better than the –5 which is the average score for a fast female. Even if a fast female decided to leave her young after being deserted by a philanderer, she would still have paid the considerable cost of an egg. So, coy genes start to spread through the population again.

To complete the hypothetical cycle, when coy females increase in numbers so much that they predominate, the philanderer males, who had such an easy time with the fast females, start to feel the pinch. Female after female insists on a long and arduous courtship. The philanderers flit from female to female, and always the story is the same. The net pay-off for a philanderer male when all the females are coy is zero. Now if a single faithful male should turn up, he is the only one with whom the coy females will mate. His net pay-off is +2, better than that of the philanderers. So, faithful genes start to increase, and we come full circle. {153}

As in the case of the aggression analysis, I have told the story as though it was an endless oscillation. But, as in that case, it can be shown that really there would be no oscillation. The system would converge to a stable state. (4) If you do the sums, it turns out that a population in which 5/6 of the females are coy, and 5/8 of the males are faithful, is evolutionarily stable. This is, of course, just for the particular arbitrary numbers that we started out with, but it is easy to work out what the stable ratios would be for any other arbitrary assumptions.

As in Maynard Smith's analyses, we do not have to think of there being two different sorts of male and two different sorts of female. The ESS could equally well be achieved if each male spends 5/8 of his time being faithful and the rest of his time philandering; and each female spends 5/6 of her time being coy and 1/6 of her time being fast. Whichever way we think of the ESS, what it means is this. Any tendency for members of either sex to deviate from their appropriate stable ratio will be penalized by a consequent change in the ratio of strategies of the other sex, which is, in turn, to the disadvantage of the original deviant. Therefore the ESS will be preserved.

We can conclude that it is certainly possible for a population consisting largely of coy females and faithful males to evolve. In these circumstances the domestic-bliss strategy for females really does seem to work. We do not have to think in terms of a conspiracy of coy females. Coyness can actually pay a female's selfish genes.

There are various ways in which females can put this type of strategy into practice. I have already suggested that a female might refuse to copulate with a male who has not already built her a nest, or at least helped her to build a nest. It is indeed the case that in many monogamous birds copulation does not take place until after the nest is built. The effect of this is that at the moment of conception the male has invested a good deal more in the child than just his cheap sperms.

Demanding that a prospective mate should build a nest is one effective way for a female to trap him. It might be thought that almost anything that costs the male a great deal would do in theory, even if that cost is not directly paid in the form of benefit to the unborn children. If all females of a population forced males to do some difficult and costly deed, like slaying a dragon or climbing a mountain, before they would consent to copulate with them, they could in theory be reducing the temptation for the males to desert {154} after copulation. Any male tempted to desert his mate and try to spread more of his genes by another female, would be put off by the thought that he would have to kill another dragon. In practice, however, it is unlikely that females would impose such arbitrary tasks as dragon-killing, or Holy-Grail-seeking on their suitors. The reason is that a rival female who imposed a task no less arduous, but more useful to her and her children, would have an advantage over more romantically minded females who demanded a pointless labour of love. Building a nest may be less romantic than slaying a dragon or swimming the Hellespont, but it is much more useful.

Also useful to the female is the practice I have already mentioned of courtship feeding by the male. In birds this has usually been regarded as a kind of regression to juvenile behaviour on the part of the female. She begs from the male, using the same gestures as a young bird would use. It has been supposed that this is automatically attractive to the male, in the same way as a man finds a lisp or pouting lips attractive in an adult woman. The female bird at this time needs all the extra food she can get, for she is building up her reserves for the effort of manufacturing her enormous eggs. Courtship feeding by the male probably represents direct investment by him in the eggs themselves. It therefore has the effect of reducing the disparity between the two parents in their initial investment in the young.

Several insects and spiders also demonstrate the phenomenon of courtship feeding. Here an alternative interpretation has sometimes been only too obvious. Since, as in the case of the praying mantis, the male may be in danger of being eaten by the larger female, anything that he can do to reduce her appetite may be to his advantage. There is a macabre sense in which the unfortunate male mantis can be said to invest in his children. He is used as food to help make the eggs which will then be fertilized, posthumously, by his own stored sperms.

A female, playing the domestic-bliss strategy, who simply looks the males over and tries to *recognize* qualities of fidelity in advance, lays herself open to deception. Any male who can pass himself off as a good loyal domestic type, but who in reality is concealing a strong tendency towards desertion and unfaithfulness, could have a great advantage. As long as his deserted former wives have any chance of bringing up some of the children, the philanderer stands to pass on more genes than a rival male who is an honest husband and father. Genes for effective deception by males will tend to be favoured in the gene pool. {155}

Conversely, natural selection will tend to favour females who become good at seeing through such deception. One way they can do this is to play especially hard to get when they are courted by a new male, but in successive breeding seasons to be increasingly ready to accept quickly the advances of last year's mate. This will automatically penalize young males embarking on their first breeding season, whether they are deceivers or not. The brood of naive first year females would tend to contain a relatively high proportion of genes from unfaithful fathers, but faithful fathers have the advantage in the second and subsequent years of a mother's life, for they do not have to go through the same prolonged energy-wasting and time-consuming courtship rituals. If the majority of individuals in a population are the children of experienced rather than naive mothers — a reasonable assumption in any long-lived species — genes for honest, good fatherhood will come to prevail in the gene pool.

For simplicity, I have talked as though a male were either purely honest or thoroughly deceitful. In reality it is more probable that all males, indeed all individuals, are a little bit deceitful, in that they are programmed to take advantage of opportunities to exploit their mates. Natural selection, by sharpening up the ability of each partner to detect dishonesty in the other, has kept large-scale deceit down to a fairly low level. Males have more to gain from dishonesty than females, and we must expect that, even in those species where males show considerable parental altruism, they will usually tend to do a bit less work than the females, and to be a bit more ready to abscond. In birds and mammals this is certainly normally the case.

There are species, however, in which the male actually does more work in caring for the children than the female does. Among birds and mammals these cases of paternal devotion are exceptionally rare, but they are common among fish. Why?(5) This is a challenge for the selfish gene theory which has puzzled me for a long time. An ingenious solution was recently suggested to me in a tutorial by Miss T. R. Carlisle. She makes use of Trivers's ‘cruel bind’ idea, referred to above, as follows.

Many fish do not copulate, but instead simply spew out their sex cells into the water. Fertilization takes place in the open water, not inside the body of one *of the* partners. This is probably how sexual reproduction first began. Land animals like birds, mammals and reptiles, on the other hand, cannot afford this kind of external {156} fertilization, because their sex cells are too vulnerable to drying-up. The gametes of one sex — the male, since sperms are mobile — are introduced into the wet interior of a member of the other sex — the female. So much is just fact. Now comes the idea. After copulation, the land-dwelling female is left in physical possession of the embryo. It is inside her body. Even if she lays the fertilized egg almost immediately, the male still has time to vanish, thereby forcing the female into Trivers's ‘cruel bind’. The male is inevitably provided with an opportunity to take the prior decision to desert, closing the female's options, and forcing her to decide whether to leave the young to certain death, or whether to stay with it and rear it. Therefore, maternal care is more common among land animals than paternal care.

But for fish and other water-dwelling animals things are very different. If the male does not physically introduce his sperms into the female's body there is no necessary sense in which the female is left ‘holding the baby’. Either partner might make a quick getaway and leave the other one in possession of the newly fertilized eggs. But there is even a possible reason why it might often be the male who is most vulnerable to being deserted. It seems probable that an evolutionary battle will develop over who sheds their sex cells first. The partner who does so has the advantage that he or she can then leave the other one in possession of the new embryos. On the other hand, the partner who spawns first runs the risk that his prospective partner may subsequently fail to follow suit. Now the male is more vulnerable here, if only because sperms are lighter and more likely to diffuse than eggs. If a female spawns too early, i.e. before the male is ready, it will not greatly matter because the eggs, being relatively large and heavy, are likely to stay together as a coherent clutch for some time. Therefore a female fish can afford to take the ‘risk’ of spawning early. The male dare not take this risk, since if he spawns too early his sperms will have diffused away before the female is ready, and she will then not spawn herself, because it will not be worth her while to do so. Because of the diffusion problem, the male must wait until the female spawns, and then he must shed his sperms over the eggs. But she has had a precious few seconds in which to disappear, leaving the male in possession, and forcing him on to the horns of Trivers's dilemma. So this theory neatly explains why paternal care is common in water but rare on dry land.

Leaving fish, I now turn to the other main female strategy, the he-man {157} strategy. In species where this policy is adopted the females, in effect, resign themselves to getting no help from the father of their children, and *go* all-out for good genes instead. Once again they use their weapon of withholding copulation. They refuse to mate with just any male, but exercise the utmost care and discrimination before they will allow a male to copulate with them. Some males undoubtedly do contain a larger number of good genes than other males, genes that would benefit the survival prospects of both sons and daughters. If a female can somehow detect good genes in males, using externally visible clues, she can benefit her own genes by allying them with good paternal genes. To use our analogy of the rowing crews, a female can minimize the chance that her genes will be dragged down through getting into bad company. She can try to hand-pick good crew-mates for her own genes.

The chances are that most of the females will agree with each other on which are the best males, since they all have the same information to go on. Therefore these few lucky males will do most of the copulating. This they are quite capable of doing, since all they must give to each female is some cheap sperms. This is presumably what has happened in elephant seals and in birds of paradise. The females are allowing just a few males to get away with the ideal selfish-exploitation strategy which all males aspire to, but they are making sure that only the best males are allowed this luxury.

From the point of view of a female trying to pick good genes with which to ally her own, what is she looking for? One thing she wants is evidence of ability to survive. Obviously any potential mate who is courting her has proved his ability to survive at least into adulthood, but he has not necessarily proved that he can survive much longer. Quite a good policy for a female might be to go for old men. Whatever their shortcomings, they have at least proved they can survive, and she is likely to be allying her genes with genes for longevity. However, there is no point in ensuring that her children live long lives if they do not also give her lots of grandchildren. Longevity is not prima facie evidence of virility. Indeed a long-lived male may have survived precisely *because* he does not take risks in order to reproduce. A female who selects an old male is not necessarily going to have more descendants than a rival female who chooses a young one who shows some other evidence of good genes.

What other evidence? There are many possibilities. Perhaps strong muscles as evidence of ability to catch food, perhaps long legs {158} as evidence of ability to run away from predators. A female might benefit her genes by allying them with such traits, since they might be useful qualities in both her sons and her daughters. To begin with, then, we have to imagine females choosing males on the basis of perfectly genuine labels or indicators which tend to be evidence of good underlying genes. But now here is a very interesting point realized by Darwin, and clearly enunciated by Fisher. In a society where males compete with each other to be chosen as he-men by females, one of the best things a mother can do for her genes is to make a son who will turn out in his turn to be an attractive he-man. If she can ensure that her son is one of the fortunate few males who wins most of the copulations in the society when he grows up, she will have an enormous number of grandchildren. The result of this is that one of the most desirable qualities a male can have in the eyes of a female is, quite simply, sexual attractiveness itself. A female who mates with a super-attractive he-man is more likely to have sons who are attractive to females of the next generation, and who will make lots of grandchildren for her. Originally, then, females may be thought of as selecting males on the basis of obviously useful qualities like big muscles, but once such qualities became widely accepted as attractive among the females of the species, natural selection would continue to favour them simply because they were attractive. Extravagances such as the tails of male birds of paradise may therefore have evolved by a kind of unstable, runaway process.(6) In the early days, a slightly longer tail than usual may have been selected by females as a desirable quality in males, perhaps because it betokened a fit and healthy constitution. A short tail on a male might have been an indicator of some vitamin deficiency — evidence of poor food-getting ability. Or perhaps short-tailed males were not very good at running away from predators, and so had had their tails bitten off. Notice that we don't have to assume that the short tail was in itself genetically inherited, only that it served as an indicator of some genetic inferiority. Anyway, for whatever reason, let us suppose that females in the ancestral bird of paradise species preferentially went for males with longer than average tails. Provided there was *some* genetic contribution to the natural variation in male tail-length, this would in time cause the average tail-length of males in the population to increase. Females followed a simple rule: look all the males over, and go for the one with the longest tail. Any female who departed from this rule was penalized, *even if* tails had already {159} become so long that they actually encumbered males possessing them. This was because any female who did not produce long-tailed sons had little chance of one of her sons being regarded as attractive. Like a fashion in women's clothes, or in American car design, the trend toward longer tails took off and gathered its own momentum. It was stopped only when tails became so grotesquely long that their manifest disadvantages started to outweigh the advantage of sexual attractiveness.

This is a hard idea to swallow, and it has attracted its sceptics ever since Darwin first proposed it, under the name of sexual selection. One person who does not believe it is A. Zahavi, whose ‘Fox, fox’ theory we have already met. He puts forward his own maddeningly contrary ‘handicap principle’ as a rival explanation.(7) He points out that the very fact that females are trying to select for good genes among males opens the door to deception by the males. Strong muscles may be a genuinely good quality for a female to select, but then what is to stop males from growing dummy muscles with no more real substance than human padded shoulders? If it costs a male less to grow false muscles than real ones, sexual selection should favour genes for producing false muscles. It will not be long, however, before counter-selection leads to the evolution of females capable of seeing through the deception. Zahavi's basic premise is that false sexual advertisement will eventually be seen through by females. He therefore concludes that really successful males will be those who do not advertise falsely, those who palpably demonstrate that they are not deceiving. If it is strong muscles we are talking about, then males who merely assume the visual *appearance* of strong muscles will soon be detected by the females. But a male who demonstrates, by the equivalent of lifting weights or ostentatiously doing press-ups, that he really has strong muscles, will succeed in convincing the females. In other words Zahavi believes that a he-man must not only *seem* to be a good quality male: he must really *be* a good quality male, otherwise he will not be accepted as such by sceptical females. Displays will therefore evolve that only a genuine he-man is capable of doing.

So far so good. Now comes the part of Zahavi's theory that really sticks in the throat. He suggests that the talis of birds of paradise and peacocks, the huge antlers of deer, and the other sexually-selected features which have always seemed paradoxical because they appear to be handicaps to their possessors, evolve precisely *because* they are {160} handicaps. A male bird with a long and cumbersome tail is showing off to females that he is such a strong he-man that he can survive *in spite of his* tail. Think of a woman watching two men run a race. If both arrive at the finishing post at the same time, but one has deliberately encumbered himself with a sack of coal on his back, the women will naturally draw the conclusion that the man with the burden is really the faster runner.

I do not believe this theory, although I am not quite so confident in my scepticism as I was when I first heard it. I pointed out then that the logical conclusion to it should be the evolution of males with only one leg and only one eye. Zahavi, who comes from Israel, instantly retorted: ‘Some of our best generals have only one eye!’ Nevertheless, the problem remains that the handicap theory seems to contain a basic contradiction. If the handicap is a genuine one — and it is of the essence of the theory that it has to be a genuine one — then the handicap itself will penalize the offspring just as surely as it may attract females. It is, in any case, important that the handicap must not be passed on to daughters.

If we rephrase the handicap theory in terms of genes, we have something like this. A gene that makes males develop a handicap, such as a long tail, becomes more numerous in the gene pool because females choose males who have handicaps. Females choose males who have handicaps, because genes that make females so choose also become frequent in the gene pool. This is because females with a taste for handicapped males will automatically tend to be selecting males with good genes in other respects, since those males have survived to adulthood in spite of the handicap. These good ‘other’ genes will benefit the bodies of the children, which therefore survive to propagate the genes for the handicap itself, and also the genes for choosing handicapped males. Provided the genes for the handicap itself exert their effect only in sons, just as the genes for a sexual preference for the handicap affect only daughters, the theory just might be made to work. So long as it is formulated only in words, we cannot be sure whether it will work or not. We get a better idea of how feasible such a theory is when it is rephrased in terms of a mathematical model. So far mathematical geneticists who have tried to make the handicap principle into a workable model have failed. This may be because it is not a workable principle, or it may be because they are not clever enough. One of them is Maynard Smith, and my hunch favours the former possibility. {161}

If a male can demonstrate his superiority over other males in a way that does not involve deliberately handicapping himself, nobody would doubt that he could increase his genetic success in that way. Thus elephant seals win and hold on to their harems, not by being aesthetically attractive to females, but by the simple expedient of beating up any male who tries to move in on the harem. Harem holders tend to win these fights against would-be usurpers, if only for the obvious reason that that is why they are harem-holders. Usurpers do not often win fights, because if they were capable of winning they would have done so before! Any female who mates only with a harem holder is therefore allying her genes with a male who is strong enough to beat off successive challenges from the large surplus of desperate bachelor males. With luck her sons will inherit their father's ability to hold a harem. In practice a female elephant seal does not have much option, because the harem-owner beats *her* up if she tries to stray. The principle remains, however, that females who choose to mate with males who win fights may benefit their genes by so doing. As we have seen, there are examples of females preferring to mate with males who hold territories and with males who have high status in the dominance hierarchy.

To sum up this chapter so far, the various different kinds of breeding system that we find among animals — monogamy, promiscuity, harems, and so on — can be understood in terms of conflicting interests between males and females. Individuals of either sex ‘want’ to maximize their total reproductive output during their lives. Because of a fundamental difference between the size and numbers of sperms and eggs, males are in general likely to be biased towards promiscuity and lack of paternal care. Females have two main available counter-ploys, which I have called the he-man and the domestic-bliss strategies. The ecological circumstances of a species will determine whether the females are biased towards one or the other of these counter-ploys, and will also determine how the males respond. In practice all intermediates between he-man and domestic-bliss are found and, as we have seen, there are cases in which the father does even more child-care than the mother. This book is not concerned with the details of particular animals species, so I will not discuss what might predispose a species towards one form of breeding system rather than another. Instead I will consider the differences that are commonly observed between males and females in general, and show how these may be interpreted. I shall therefore not be emphasizing those species in which the differences between the sexes are slight, these being in general the ones whose females have favoured the domestic-bliss strategy.

Firstly, it tends to be the males who go in for sexually attractive, gaudy colours, and the females who tend to be more drab. Individuals of both sexes want to avoid being eaten by predators, and there will be some evolutionary pressure on both sexes to be drably coloured. Bright colours attract predators no less than they attract sexual partners. In gene terms, this means that genes for bright colours are more likely to meet their end in the stomachs of predators than are genes for drab colours. On the other hand, genes for drab colours may be less likely than genes for bright colours to find themselves in the next generation, because drab individuals have difficulty in attracting a mate. There are therefore two conflicting selection pressures: predators tending to remove bright-colour genes from the gene pool, and sexual partners tending to remove genes for drabness. As in so many other cases, efficient survival machines can be regarded as a compromise between conflicting selection pressures. What interests us at the moment is that the optimal compromise for a male seems to be different from the optimal compromise for a female. This is of course fully compatible with our view of males as high-risk, high-reward gamblers. Because a male produces many millions of sperms to every egg produced by a female, sperms heavily outnumber eggs in the population. Any given egg is therefore much more likely to enter into sexual fusion than any given sperm is. Eggs are a relatively valuable resource, and therefore a female does not need to be so sexually attractive as a male does in order to ensure that her eggs are fertilized. A male is perfectly capable of siring all the children bom to a large population of females. Even if a male has a short life because his gaudy tail attracts predators, or gets tangled in the bushes, he may have fathered a very large number of children before he dies. An unattractive or drab male may live even as long as a female, but he has few children, and his genes are not passed on. What shall it profit a male if he shall gain the whole world, and lose his immortal genes?

Another common sexual difference is that females are more fussy than males about whom they mate with. One of the reasons for fussiness by an individual of either sex is the need to avoid mating with a member of another species. Such hybridizations are a bad thing for a variety of reasons. Sometimes, as in the case of a man {163} copulating with a sheep, the copulation does not lead to an embryo being formed, so not much is lost. When more closely related species like horses and donkeys cross-breed, however, the cost, at least to the female partner, can be considerable. An embryo mule is likely to be formed and it then clutters up her womb for eleven months. It takes a large quantity of her total parental investment, not only in the form of food absorbed through the placenta, and then later in the form of milk, but above all in time which could have been spent in rearing other children. Then when the mule reaches adulthood it turns out to be sterile. This is presumably because, although horse chromosomes and donkey chromosomes are sufficiently similar to cooperate in the building of a good strong mule body, they are not similar enough to work together properly in meiosis. Whatever the exact reason, the very considerable investment by the mother in the rearing of a mule is totally wasted from the point of view of her genes. Female horses should be very, very careful that the individual they copulate with is another horse, and not a donkey. In gene terms, any horse gene that says ‘Body, if you are female, copulate with any old male, whether he is a donkey or a horse’, is a gene which may next find itself in the dead-end body of a mule, and the mother's parental investment in that baby mule detracts heavily from her capacity to rear fertile horses. A male, on the other hand, has less to lose if he mates with a member of the wrong species, and, although he may have nothing to gain either, we should expect males to be less fussy in their choice of sexual partners. Where this has been looked at, it has been found to be true.

Even within a species, there may be reasons for fussiness. Incestuous mating, like hybridization, is likely to have damaging genetic consequences, in this case because lethal and semi-lethal recessive genes are brought out into the open. Once again, females have more to lose than males, since their investment in any particular child tends to be greater. Where incest taboos exist, we should expect females to be more rigid in their adherence to the taboos than males. If we assume that the older partner in an incestuous relationship is relatively likely to be the active initiator, we should expect that incestuous unions in which the male is older than the female should be more common than unions in which the female is older.

For instance father/daughter incest should be commoner than mother/son. Brother/sister incest should be intermediate in commonness.

In general, males should tend to be more promiscuous than {164} females. Since a female produces a limited number of eggs at a relatively slow rate, she has little to gain from having a large number of copulations with different males. A male on the other hand, who can produce millions of sperms every day, has everything to gain from as many promiscuous matings as he can snatch. Excess copulations may not actually cost a female much, other than a little lost time and energy, but they do not do her positive good. A male on the other hand can never get enough copulations with as many different females as possible: the word excess has no meaning for a male.

I have not explicitly talked about man but inevitably, when we think about evolutionary arguments such as those in this chapter, we cannot help reflecting about our own species and our own experience. Notions of females withholding copulation until a male shows some evidence of long-term fidelity may strike a familiar chord. This might suggest that human females play the domestic-bliss rather than the he-man strategy. Many human societies are indeed monogamous. In our own society, parental investment by both parents is large and not obviously unbalanced. Mothers certainly do more direct work for children than fathers do, but fathers often work hard in a more indirect sense to provide the material resources that are poured into the children. On the other hand, some human societies are promiscuous, and many are harem-based. What this astonishing variety suggests is that man's way of life is largely determined by culture rather than by genes. However, it is still possible that human males in general have a tendency towards promiscuity, and females a tendency towards monogamy, as we would predict on evolutionary grounds. Which of these two tendencies wins in particular societies depends on details of cultural circumstance, just as in different animal species it depends on ecological details.

One feature of our own society that seems decidedly anomalous is the matter of sexual advertisement. As we have seen, it is strongly to be expected on evolutionary grounds that, where the sexes differ, it should be the males that advertise and the females that are drab. Modern western man is undoubtedly exceptional in this respect. It is of course true that some men dress flamboyantly and some women dress drably but, on average, there can be no doubt that in our society the equivalent of the peacock's tail is exhibited by the female, not by the male. Women paint their faces and glue on false eyelashes. Apart from special cases, like actors, men do not. Women seem to be {165} interested in their own personal appearance and they are encouraged in this by their magazines and journals. Men's magazines are less preoccupied with male sexual attractiveness, and a man who is unusually interested in his own dress and appearance is apt to arouse suspicion, both among men and among women. When a woman is described in conversation, it is quite likely that her sexual attractiveness, or lack of it, will be prominently mentioned. This is true, whether the speaker is a man or a woman. When a man is described, the adjectives used are much more likely to have nothing to do with sex.

Faced with these facts, a biologist would be forced to suspect that he was looking at a society in which females compete for males, rather than vice versa. In the case of birds of paradise, we decided that females are drab because they do not need to compete for males. Males are bright and ostentatious because females are in demand and can afford to be choosy. The reason female birds of paradise are in demand is that eggs are a more scarce resource than sperms. What has happened in modern western man? Has the male really become the sought-after sex, the one that is in demand, the sex that can afford to be choosy? If so, why?

11 MEMES: THE NEW REPLICATORS

So far, I have not talked much about man in particular, though I have not deliberately excluded him either. Part of the reason I have used the term ‘survival machine’ is that ‘animal’ would have left out plants and, in some people's minds, humans. The arguments I have put forward should, prima facie, apply to any evolved being. If a species is to be excepted, it must be for good particular reasons. Are there any good reasons for supposing our own species to be unique? I believe the answer is yes.

Most of what is unusual about man can be summed up in one word: ‘culture’. I use the word not in its snobbish sense, but as a scientist uses it. Cultural transmission is analogous to genetic transmission in that, although basically conservative, it can give rise to a form of evolution. Geoffrey Chaucer could not hold a conversation with a modern Englishman, even though they are linked to each other by an unbroken chain of some twenty generations of Englishmen, each of whom could speak to his immediate neighbours in the chain as a son speaks to his father. Language seems to ‘evolve’ by non-genetic means, and at a rate which is orders of magnitude faster than genetic evolution.

Cultural transmission is not unique to man. The best non-human example that I know has recently been described by P. F. Jenkins in the song of a bird called the saddleback which lives on islands off New Zealand. On the island where he worked there was a total repertoire of about nine distinct songs. Any given male sang only one or a few of these songs. The males could be classified into dialect groups. For example, one group of eight males with neighbouring territories sang a particular song called the CC song. Other dialect groups sang different songs. Sometimes the members of a dialect group shared more than one distinct song. By comparing the songs of fathers and sons, Jenkins showed that song patterns were not inherited genetically. Each young male was likely to adopt songs {190} from his territorial neighbours by imitation, in an analogous way to human language. During most of the time Jenkins was there, there was a fixed number of songs on the island, a kind of ‘song pool’ from which each young male drew his own small repertoire. But occasionally Jenkins was privileged to witness the ‘invention’ of a new song, which occurred by a mistake in the imitation of an old one. He writes: ‘New song forms have been shown to arise variously by change of pitch of a note, repetition of a note, the elision of notes and the combination of parts of other existing songs . . . The appearance of the new form was an abrupt event and the product was quite stable over a period of years. Further, in a number of cases the variant was transmitted accurately in its new form to younger recruits so that a recognizably coherent group of like singers developed.’ Jenkins refers to the origins of new songs as ‘cultural mutations’.

Song in the saddleback truly evolves by non-genetic means. There are other examples of cultural evolution in birds and monkeys, but these are just interesting oddities. It is our own species that really shows what cultural evolution can do. Language is only one example out of many. Fashions in dress and diet, ceremonies and customs, art and architecture, engineering and technology, all evolve in historical time in a way that looks like highly speeded up genetic evolution, but has really nothing to do with genetic evolution. As in genetic evolution though, the change may be progressive. There is a sense in which modern science is actually better than ancient science. Not only does our understanding of the universe change as the centuries go by: it improves. Admittedly the current burst of improvement dates back only to the Renaissance, which was preceded by a dismal period of stagnation, in which European scientific culture was frozen at the level achieved by the Greeks. But, as we saw in Chapter 5, genetic evolution too may proceed as a series of brief spurts between stable plateaux.

The analogy between cultural and genetic evolution has frequently been pointed out, sometimes in the context of quite unnecessary mystical overtones. The analogy between scientific progress and genetic evolution by natural selection has been illuminated especially by Sir Karl Popper. I want to go even further into directions which are also being explored by, for example, the geneticist L. L. Cavalli-Sforza, the anthropologist F. T. Cloak, and the ethologist J. M. Cullen.

As an enthusiastic Darwinian, I have been dissatisfied with {191} explanations that my fellow-enthusiasts have offered for human behaviour. They have tried to look for ‘biological advantages’ in various attributes of human civilization. For instance, tribal religion has been seen as a mechanism for solidifying group identity, valuable for a pack-hunting species whose individuals rely on cooperation to catch large and fast prey. Frequently the evolutionary preconception in terms of which such theories are framed is implicitly group-selectionist, but it is possible to rephrase the theories in terms of orthodox gene selection. Man may well have spent large portions of the last several million years living in small kin groups. Kin selection and selection in favour of reciprocal altruism may have acted on human genes to produce many of our basic psychological attributes and tendencies. These ideas are plausible as far as they go, but I find that they do not begin to square up to the formidable challenge of explaining culture, cultural evolution, and the immense differences between human cultures around the world, from the utter selfishness of the Ik of Uganda, as described by Colin Turnbull, to the gentle altruism of Margaret Mead's Arapesh. I think we have got to start again and go right back to first principles. The argument I shall advance, surprising as it may seem coming from the author of the earlier chapters, is that, for an understanding of the evolution of modern man, we must begin by throwing out the gene as the sole basis of our ideas on evolution. I am an enthusiastic Darwinian, but I think Darwinism is too big a theory to be confined to the narrow context of the gene. The gene will enter my thesis as an analogy, nothing more.

What, after all, is so special about genes? The answer is that they are replicators. The laws of physics are supposed to be true all over the accessible universe. Are there any principles of biology that are likely to have similar universal validity? When astronauts voyage to distant planets and look for life, they can expect to find creatures too strange and unearthly for us to imagine. But is there anything that must be true of all life, wherever it is found, and whatever the basis of its chemistry? If forms of life exist whose chemistry is based on silicon rather than carbon, or ammonia rather than water, if creatures are discovered that boil to death at –100 degrees centigrade, if a form of life is found that is not based on chemistry at all but on electronic reverberating circuits, will there still be any general principle that is true of all life? Obviously I do not know but, if I had to bet, I would put my money on one fundamental principle. This is {192} the law that all life evolves by the differential survival of replicating entities.(1) The gene, the DNA molecule, happens to be the replicating entity that prevails on our own planet. There may be others. If there are, provided certain other conditions are met, they will almost inevitably tend to become the basis for an evolutionary process.

But do we have to go to distant worlds to find other kinds of replicator and other, consequent, kinds of evolution? I think that a new kind of replicator has recently emerged on this very planet. It is staring us in the face. It is still in its infancy, still drifting clumsily about in its primeval soup, but already it is achieving evolutionary change at a rate that leaves the old gene panting far behind.

The new soup is the soup of human culture. We need a name for the new replicator, a noun that conveys the idea of a unit of cultural transmission, or a unit of *imitation*. ‘Mimeme’ comes from a suitable Greek root, but I want a monosyllable that sounds a bit like ‘gene’. I hope my classicist friends will forgive me if I abbreviate mimeme to *meme*(2) If it is any consolation, it could alternatively be thought of as being related to ‘memory’, or to the French word *meme*. It should be pronounced to rhyme with ‘cream’.

Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation. If a scientist hears, or reads about, a good idea, he passes it on to his colleagues and students. He mentions it in his articles and his lectures. If the idea catches on, it can be said to propagate itself, spreading from brain to brain. As my colleague N. K. Humphrey neatly summed up an earlier draft of this chapter: ‘. . . memes should be regarded as living structures, not just metaphorically but technically.(3) When you plant a fertile meme in my mind you literally parasitize my brain, turning it into a vehicle for the meme's propagation in just the way that a virus may parasitize the genetic mechanism of a host cell. And this isn't just a way of talking — the meme for, say, "belief in life after death" is actually realized physically, millions of times over, as a structure in the nervous systems of individual men the world over.’

Consider the idea of God. We do not know how it arose in the meme pool. Probably it originated many times by independent ‘mutation’. In any case, it is very old indeed. How does it replicate {193} itself? By the spoken and written word, aided by great music and great art. Why does it have such high survival value? Remember that ‘survival value’ here does not mean value for a gene in a gene pool, but value for a meme in a meme pool. The question really means: What is it about the idea of a god that gives it its stability and penetrance in the cultural environment? The survival value of the god meme in the meme pool results from its great psychological appeal. It provides a superficially plausible answer to deep and troubling questions about existence. It suggests that injustices in this world may be rectified in the next. The ‘everlasting arms’ hold out a cushion against our own inadequacies which, like a doctor's placebo, is none the less effective for being imaginary. These are some of the reasons why the idea of God is copied so readily by successive generations of individual brains. God exists, if only in the form of a meme with high survival value, or infective power, in the environment provided by human culture.

Some of my colleagues have suggested to me that this account of the survival value of the god meme begs the question. In the last analysis they wish always to go back to ‘biological advantage’. To them it is not good enough to say that the idea of a god has ‘great psychological appeal’. They want to know *why* it has great psychological appeal. Psychological appeal means appeal to brains, and brains are shaped by natural selection of genes in gene-pools. They want to find some way in which having a brain like that improves gene survival.

I have a lot of sympathy with this attitude, and I do not doubt that there are genetic advantages in our having brains of the kind that we have. But nevertheless I think that these colleagues, if they look carefully at the fundamentals of their own assumptions, will find that they are begging just as many questions as I am. Fundamentally, the reason why it is good policy for us to try to explain biological phenomena in terms of gene advantage is that genes are replicators. As soon as the primeval soup provided conditions in which molecules could make copies of themselves, the replicators themselves took over. For more than three thousand million years, DNA has been the only replicator worth talking about in the world. But it does not necessarily hold these monopoly rights for all time. Whenever conditions arise in which a new kind of replicator *can* make copies of itself, the new replicators *will* tend to take over, and start a new kind of evolution of their own. Once this new evolution begins, it will in no {194} necessary sense be subservient to the old. The old gene-selected evolution, by making brains, provided the soup' in which the first memes arose. Once self-copying memes had arisen, their own, much faster, kind of evolution took off. We biologists have assimilated the idea of genetic evolution so deeply that we tend to forget that it is only one of many possible kinds of evolution.

Imitation, in the broad sense, is how memes *can* replicate. But just as not all genes that can replicate do so successfully, so some memes are more successful in the meme-pool than others. This is the analogue of natural selection. I have mentioned particular examples of qualities that make for high survival value among memes. But in general they must be the same as those discussed for the replicators of Chapter 2: longevity, fecundity, and copying-fidelity. The longevity of any one copy of a meme is probably relatively unimportant, as it is for any one copy of a gene. The copy of the tune 'Auld Lang Syne' that exists in my brain will last only for the rest of my life.(4) The copy of the same tune that is printed in my volume of *The Scottish Student's Song Book* is unlikely to last much longer. But I expect there will be copies of the same tune on paper and in peoples' brains for centuries to come. As in the case of genes, fecundity is much more important than longevity of particular copies. If the meme is a scientific idea, its spread will depend on how acceptable it is to the population of individual scientists; a rough measure of its survival value could be obtained by counting the number of times it is referred to in successive years in scientific journals.(5) If it is a popular tune, its spread through the meme pool may be gauged by the number of people heard whistling it in the streets. If it is a style of women's shoe, the population memeticist may use sales statistics from shoe shops. Some memes, like some genes, achieve brilliant short-term success in spreading rapidly, but do not last long in the meme pool. Popular songs and stiletto heels are examples. Others, such as the Jewish religious laws, may continue to propagate themselves for thousands of years, usually because of the great potential permanence of written records.

This brings me to the third general quality of successful replicators: copying-fidelity. Here I must admit that I am on shaky ground. At first sight it looks as if memes are not high-fidelity replicators at all. Every time a scientist hears an idea and passes it on to somebody else, he is likely to change it somewhat. I have made no secret of my debt in this book to the ideas of R. L. Trivers. Yet I have {195} not repeated them in his own words. I have twisted them round for my own purposes, changing the emphasis, blending them with ideas of my own and of other people. The memes are being passed on to you in altered form. This looks quite unlike the particulate, all-or-none quality of gene transmission. It looks as though meme transmission is subject to continuous mutation, and also to blending.

It is possible that this appearance of non-particulateness is illusory, and that the analogy with genes does not break down. After all, if we look at the inheritance of many genetic characters such as human height or skin-colouring, it does not look like the work of indivisible and unblendable genes. If a black and a white person mate, their children do not come out either black or white: they are intermediate. This does not mean the genes concerned are not particulate. It is just that there are so many of them concerned with skin colour, each one having such a small effect, that they seem to blend. So far I have talked of memes as though it was obvious what a single unit-meme consisted of. But of course it is far from obvious. I have said a tune is one meme, but what about a symphony: how many memes is that? Is each movement one meme, each recognizable phrase of melody, each bar, each chord, or what?

I appeal to the same verbal trick as I used in Chapter 3. There I divided the ‘gene complex’ into large and small genetic units, and units within units. The ‘gene’ was defined, not in a rigid all-or-none way, but as a unit of convenience, a length of chromosome with just sufficient copying-fidelity to serve as a viable unit of natural selection. If a single phrase of Beethoven's ninth symphony is sufficiently distinctive and memorable to be abstracted from the context of the whole symphony, and used as the call-sign of a maddeningly intrusive European broadcasting station, then to that extent it deserves to be called one meme. It has, incidentally, materially diminished my capacity to enjoy the original symphony.

Similarly, when we say that all biologists nowadays believe in Darwin's theory, we do not mean that every biologist has, graven in his brain, an identical copy of the exact words of Charles Darwin himself. Each individual has his own way of interpreting Darwin's ideas. He probably learned them not from Darwin's own writings, but from more recent authors. Much of what Darwin said is, in detail, wrong. Darwin if he read this book would scarcely recognize his own original theory in it, though I hope he would like the way I put it. Yet, in spite of all this, there is something, some essence of {196} Darwinism, which is present in the head of every individual who understands the theory. If this were not so, then almost any statement about two people agreeing with each other would be meaningless. An ‘idea-meme’ might be defined as an entity that is capable of being transmitted from one brain to another. The meme of Darwin's theory is therefore that essential basis of the idea which is held in common by all brains that understand the theory. The *differences* in the ways that people represent the theory are then, by definition, not part of the meme. If Darwin's theory can be sub-divided into components, such that some people believe component *A* but not component *B*, while others believe *B* but *not A*, then *A* and *B* should be regarded as separate memes. If almost everybody who believes *in A* also believes in *B* — if the memes are closely ‘linked’ to use the genetic term — then it is convenient to lump them together as one meme.

Let us pursue the analogy between memes and genes further. Throughout this book, I have emphasized that we must not think of genes as conscious, purposeful agents. Blind natural selection, however, makes them behave rather as if they were purposeful, and it has been convenient, as a shorthand, to refer to genes in the language of purpose. For example, when we say ‘genes are trying to increase their numbers in future gene pools’, what we really mean is ‘those genes that behave in such a way as to increase their numbers in future gene pools tend to be the genes whose effects we see in the world’. Just as we have found it convenient to think of genes as active agents, working purposefully for their own survival, perhaps it might be convenient to think of memes in the same way. In neither case must we get mystical about it. In both cases the idea of purpose is only a metaphor, but we have already seen what a fruitful metaphor it is in the case of genes. We have even used words like ‘selfish’ and ‘ruthless’ of genes, knowing full well it is only a figure of speech. Can we, in exactly the same spirit, look for selfish or ruthless memes?

There is a problem here concerning the nature of competition. Where there is sexual reproduction, each gene is competing particularly with its own alleles — rivals for the same chromosomal slot. Memes seem to have nothing equivalent to chromosomes, and nothing equivalent to alleles. I suppose there is a trivial sense in which many ideas can be said to have 'opposites'. But in general memes resemble the early replicating molecules, floating chaotically free in the primeval soup, rather than modern genes in their neatly {197} paired, chromosomal regiments. In what sense then are memes competing with each other? Should we expect them to be 'selfish' or 'ruthless', if they have no alleles? The answer is that we might, because there is a sense in which they must indulge in a kind of competition with each other.

Any user of a digital computer knows how precious computer time and memory storage space are. At many large computer centres they are literally costed in money; or each user may be allotted a ration of time, measured in seconds, and a ration of space, measured in 'words'. The computers in which memes live are human brains.(6) Time is possibly a more important limiting factor than storage space, and it is the subject of heavy competition. The human brain, and the body that it controls, cannot do more than one or a few things at once. If a meme is to dominate the attention of a human brain, it must do so at the expense of 'rival' memes. Other commodities for which memes compete are radio and television time, billboard space, newspaper column-inches, and library shelf-space.

In the case of genes, we saw in Chapter 3 that co-adapted gene complexes may arise in the gene pool. A large set of genes concerned with mimicry in butterflies became tightly linked together on the same chromosome, so tightly that they can be treated as one gene. In Chapter 5 we met the more sophisticated idea of the evolutionarily stable set of genes. Mutually suitable teeth, claws, guts, and sense organs evolved in carnivore gene pools, while a different stable set of characteristics emerged from herbivore gene pools. Does anything analogous occur in meme pools? Has the god meme, say, become associated with any other particular memes, and does this association assist the survival of each of the participating memes? Perhaps we could regard an organized church, with its architecture, rituals, laws, music, art, and written tradition, as a co-adapted stable set of mutually-assisting memes.

To take a particular example, an aspect of doctrine that has been very effective in enforcing religious observance is the threat of hell fire. Many children and even some adults believe that they will suffer ghastly torments after death if they do not obey the priestly rules. This is a peculiarly nasty technique of persuasion, causing great psychological anguish throughout the middle ages and even today. But it is highly effective. It might almost have been planned deliberately by a machiavellian priesthood trained in deep psychological indoctrination techniques. However, I doubt if the priests {198} were that clever. Much more probably, unconscious memes have ensured their own survival by virtue of those same qualities of pseudo-ruthlessness that successful genes display. The idea of hell fire is, quite simply, *self perpetuating*, because of its own deep psychological impact. It has become linked with the god meme because the two reinforce each other, and assist each other's survival in the meme pool.

Another member of the religious meme complex is called faith. It means blind trust, in the absence of evidence, even in the teeth of evidence. The story of Doubting Thomas is told, not so that we shall admire Thomas, but so that we can admire the other apostles in comparison. Thomas demanded evidence. Nothing is more lethal for certain kinds of meme than a tendency to look for evidence. The other apostles, whose faith was so strong that they did not need evidence, are held up to us as worthy of imitation. The meme for blind faith secures its own perpetuation by the simple unconscious expedient of discouraging rational inquiry.

Blind faith can justify anything.(7) If a man believes in a different god, or even if he uses a different ritual for worshipping the same god, blind faith can decree that he should die — on the cross, at the stake, skewered on a Crusader's sword, shot in a Beirut street, or blown up in a bar in Belfast. Memes for blind faith have their own ruthless ways of propagating themselves. This is true of patriotic and political as well as religious blind faith.

Memes and genes may often reinforce each other, but they sometimes come into opposition. For example, the habit of celibacy is presumably not inherited genetically. A gene for celibacy is doomed to failure in the gene pool, except under very special circumstances such as we find in the social insects. But still, a *meme* for celibacy can be successful in the meme pool. For example, suppose the success of a meme depends critically on how much time people spend in actively transmitting it to other people. Any time spent in doing other things than attempting to transmit the meme may be regarded as time wasted from the meme's point of view. The meme for celibacy is transmitted by priests to young boys who have not yet decided what they want to do with their lives. The medium of transmission is human influence of various kinds, the spoken and written word, personal example and so on. Suppose, for the sake of argument, it happened to be the case that marriage weakened the power of a priest to influence his flock, say because it occupied a {199} large proportion of his time and attention. This has, indeed, been advanced as an official reason for the enforcement of celibacy among priests. If this were the case, it would follow that the meme for celibacy could have greater survival value than the meme for marriage. Of course, exactly the opposite would be true for a *gene* for celibacy. If a priest is a survival machine for memes, celibacy is a useful attribute to build into him. Celibacy is just a minor partner in a large complex of mutually-assisting religious memes.

I conjecture that co-adapted meme-complexes evolve in the same kind of way as co-adapted gene-complexes. Selection favours memes that exploit their cultural environment to their own advantage. This cultural environment consists of other memes which are also being selected. The meme pool therefore comes to have the attributes of an evolutionarily stable set, which new memes find it hard to invade.

I have been a bit negative about memes, but they have their cheerful side as well. When we die there are two things we can leave behind us: genes and memes. We were built as gene machines, created to pass on our genes. But that aspect of us will be forgotten in three generations. Your child, even your grandchild, may bear a resemblance to you, perhaps in facial features, in a talent for music, in the colour of her hair. But as each generation passes, the contribution of your genes is halved. It does not take long to reach negligible proportions. Our genes may be immortal but the *collection* of genes that is any one of us is bound to crumble away. Elizabeth II is a direct descendant of William the Conqueror. Yet it is quite probable that she bears not a single one of the old king's genes. We should not seek immortality in reproduction.

But if you contribute to the world's culture, if you have a good idea, compose a tune, invent a sparking plug, write a poem, it may live on, intact, long after your genes have dissolved in the common pool. Socrates may or may not have a gene or two alive in the world today, as G. C. Williams has remarked, but who cares? The meme-complexes of Socrates, Leonardo, Copernicus and Marconi are still going strong.

However speculative my development of the theory of memes may be, there is one serious point which I would like to emphasize once again. This is that when we look at the evolution of cultural traits and at their survival value, we must be clear *whose* survival we are talking about. Biologists, as we have seen, are accustomed to looking for {200} advantages at the gene level (or the individual, the group, or the species level according to taste). What we have not previously considered is that a cultural trait may have evolved in the way that it has, simply because it is *advantageous to itself.*

We do not have to look for conventional biological survival values of traits like religion, music, and ritual dancing, though these may also be present. Once the genes have provided their survival machines with brains that are capable of rapid imitation, the memes will automatically take over. We do not even have to posit a genetic advantage in imitation, though that would certainly help. All that is necessary is that the brain should be *capable* of imitation: memes will then evolve that exploit the capability to the full.

I now close the topic of the new replicators, and end the chapter on a note of qualified hope. One unique feature of man, which may or may not have evolved memically, is his capacity for conscious foresight. Selfish genes (and, if you allow the speculation of this chapter, memes too) have no foresight. They are unconscious, blind, replicators. The fact that they replicate, together with certain further conditions means, willy nilly, that they will tend towards the evolution of qualities which, in the special sense of this book, can be called selfish. A simple replicator, whether gene or meme, cannot be expected to forgo short-term selfish advantage even if it would really pay it, in the long term, to do so. We saw this in the chapter on aggression. Even though a ‘conspiracy of doves’ would be better for *every single individual* than the evolutionarily stable strategy, natural selection is bound to favour the ESS.

It is possible that yet another unique quality of man is a capacity for genuine, disinterested, true altruism. I hope so, but I am not going to argue the case one way or the other, nor to speculate over its possible memic evolution. The point I am making now is that, even if we look on the dark side and assume that individual man is fundamentally selfish, our conscious foresight — our capacity to simulate the future in imagination — could save us from the worst selfish excesses of the blind replicators. We have at least the mental equipment to foster our long-term selfish interests rather than merely our short-term selfish interests. We can see the long-term benefits of participating in a ‘conspiracy of doves’, and we can sit down together to discuss ways of making the conspiracy work. We have the power to defy the selfish genes of our birth and, if necessary, the selfish memes of our indoctrination. We can even discuss ways of {201} deliberately cultivating and nurturing pure, disinterested altruism — something that has no place in nature, something that has never existed before in the whole history of the world. We are built as gene machines and cultured as meme machines, but we have the power to turn against our creators. We, alone on earth, can rebel against the tyranny of the selfish replicators.(8)

12 NICE GUYS FINISH FIRST

Nice guys finish last. The phrase seems to have originated in the world of baseball, although some authorities claim priority for an alternative connotation. The American biologist Garrett Hardin used it to summarize the message of what may be called ‘sociobiology’ or ‘selfish genery’. It is easy to see its aptness. If we translate the colloquial meaning of ‘nice guy’ into its Darwinian equivalent, a nice guy is an individual that assists other members of its species, at its own expense, to pass their genes on to the next generation. Nice guys, then, seem bound to decrease in numbers: niceness dies a Darwinian death. But there is another, technical, interpretation of the colloquial word ‘nice’. If we adopt this definition, which is not too far from the colloquial meaning, nice guys can *finish first*. This more optimistic conclusion is what this chapter is about.

Remember the Grudgers of Chapter 10. These were birds that helped each other in an apparently altruistic way, but refused to help — bore a grudge against — individuals that had previously refused to help them. Grudgers came to dominate the population because they passed on more genes to future generations than either Suckers (who helped others indiscriminately, and were exploited) or Cheats (who tried ruthlessly to exploit everybody and ended up doing each other down). The story of the Grudgers illustrated an important general principle, which Robert Trivers called ‘reciprocal altruism’. As we saw in the example of the cleaner fish (pages 186-7), reciprocal altruism is not confined to members of a single species. It is at work in all relationships that are called symbiotic — for instance the ants milking their aphid ‘cattle’ (page 181). Since Chapter 10 was written, the American political scientist Robert Axelrod (working partly in collaboration with W. D. Hamilton, whose name has cropped up on so many pages of this book), has taken the idea of reciprocal altruism on in exciting new directions. It was Axelrod who coined the technical meaning of the word ‘nice’ to which I alluded in my opening paragraph. {203}

Axelrod, like many political scientists, economists, mathematicians and psychologists, was fascinated by a simple gambling game called Prisoner's Dilemma. It is so simple that I have known clever men misunderstand it completely, thinking that there must be more to it! But its simplicity is deceptive. Whole shelves in libraries are devoted to the ramifications of this beguiling game. Many influential people think it holds the key to strategic defence planning, and that we should study it to prevent a third world war. As a biologist, I agree with Axelrod and Hamilton that many wild animals and plants are engaged in ceaseless games of Prisoner's Dilemma, played out in evolutionary time.

In its original, human, version, here is how the game is played. There is a ‘banker’, who adjudicates and pays out winnings to the two players. Suppose that I am playing against you (though, as we shall see, ‘against’ is precisely what we don't have to be). There are only two cards in each of our hands, labelled COOPERATE and DEFECT. To play, we each choose one of our cards and lay it face down on the table. Face down so that neither of us can be influenced by the other's move: in effect, we move simultaneously. We now wait in suspense for the banker to turn the cards over. The suspense is because our winnings depend not just on which card we have played (which we each know), but on the other player's card too (which we don't know until the banker reveals it).

Since there are 2 × 2 cards, there are four possible outcomes. For each outcome, our winnings are as follows (quoted in dollars in deference to the North American origins of the game):

Outcome I: We have both played COOPERATE. The banker pays each of us $300. This respectable sum is called the Reward for mutual cooperation.

Outcome II: We have both played DEFECT. The banker fines each of us $10. This is called the Punishment for mutual defection.

Outcome III: You have played COOPERATE; I have played DEFECT. The banker pays me $500 (the Temptation to defect) and fines you (the Sucker)

$100.

Outcome IV: You have played DEFECT; I have played COOPERATE. The banker pays you the Temptation payoff of $500 and fines me, the Sucker, $100. {204}

Outcomes III and IV are obviously mirror images: one player does very well and the other does very badly. In outcomes I and II we do as well as one another, but I is better for *both* of us than II. The exact quantities of money don't matter. It doesn't even matter how many of them are positive (payments) and how many of them, if any, are negative (fines). What matters, for the game to qualify as a true Prisoner's Dilemma, is their rank order. The Temptation to defect must be better than the Reward for mutual cooperation, which must be better than the Punishment for mutual defection, which must be better than the Sucker's payoff. (Strictly speaking, there is one further condition for the game to qualify as a true Prisoner's Dilemma: the average of the Temptation and the Sucker payoffs must not exceed the Reward. The reason for this additional condition will emerge later.) The four outcomes are summarized in the payoff matrix in Figure A.

		What you do	
		Cooperate	**Defect**
What I do	**Cooperate**	Fairly good **REWARD** (for mutual cooperation)	Very bad **SUCKER'S PAYOFF**

	e.g. $300	e.g. $100 fine
Defect	Very good **TEMPTATION** (to defect) e.g. $500	Fairly bad **PUNISHMENT** (for mutual defection) e.g. $10 fine

Figure A. Payoffs to me from various outcomes of the Prisoner's Dilemma game

Now, why the ‘dilemma’? To see this, look at the payoff matrix and imagine the thoughts that might go through my head as I play against you. I know that there are only two cards you can play, COOPERATE and DEFECT. Let's consider them in order. If you have played DEFECT (this means we have to look at the right hand column), the {205} best card I could have played would have been DEFECT too. Admittedly I'd have suffered the penalty for mutual defection, but if I'd cooperated I'd have got the Sucker's payoff which is even worse. Now let's turn to the other thing you could have done (look at the left hand column), play the COOPERATE card. Once again DEFECT is the best thing I could have done. If I had cooperated we'd both have got the rather high score of $300. But if I'd defected I'd have got even more — $500. The conclusion is that, regardless of which card you play, my best move is *Always Defect.*

So I have worked out by impeccable logic that, regardless of what you do, I must defect. And you, with no less impeccable logic, will work out just the same thing. So when two rational players meet, they will both defect, and both will end up with a fine or a low payoff. Yet each knows perfectly well that, if only they had *both* played COOPERATE, both would have obtained the relatively high reward for mutual cooperation ($300 in our example).

That is why the game is called a dilemma, why it seems so maddeningly paradoxical, and why it has even been proposed that there ought to be a law against it.

‘Prisoner’ comes from one particular imaginary example. The currency in this case is not money but prison sentences. Two men — call them Peterson and Moriarty — are in jail, suspected of collaborating in a crime. Each prisoner, in his separate cell, is invited to betray his colleague (DEFECT) by turning King's Evidence against him. What happens depends upon what both prisoners do, and neither knows what the other has done. If Peterson throws the blame entirely on Moriarty, and Moriarty renders the story plausible by remaining silent (cooperating with his erstwhile and, as it turns out, treacherous friend), Moriarty gets a heavy jail sentence while Peterson gets off scot-free, having yielded to the Temptation to defect. If each betrays the other, both are convicted of the crime, but receive some credit for giving evidence and get a somewhat reduced, though still stiff, sentence, the Punishment for mutual defection. If both cooperate (with each other, not with the authorities) by refusing to speak, there is not enough evidence to convict either of them of the main crime, and they receive a small sentence for a lesser offence, the Reward for mutual cooperation. Although it may seem odd to call a jail sentence a ‘reward’, that is how the men would see it if the alternative was a longer spell behind bars. You will notice that, although the ‘payoffs’ are not in dollars but in jail sentences, the {206} essential features of the game are preserved (look at the rank order of desirability of the four outcomes). If you put yourself in each prisoner's place, assuming both to be motivated by rational self-interest and remembering that they cannot talk to one another to make a pact, you will see that neither has any choice but to betray the other, thereby condemning both to heavy sentences.

Is there any way out of the dilemma? Both players know that, whatever their opponent does, they themselves cannot do better than DEFECT; yet both also know that, if only *both* had cooperated, *each* one would have done better. If only. . . if only. . . if only there could be some way of reaching agreement, some way of reassuring each player that the other can be trusted not to go for the selfish jackpot, some way of policing the agreement.

In the simple game of Prisoner's Dilemma, there is no way of ensuring trust. Unless at least one of the players is a really saintly sucker, too good for this world, the game is doomed to end in mutual defection with its paradoxically poor result for both players. But there is another version of the game. It is called the ‘Iterated’ or ‘Repeated’ Prisoner's Dilemma. The iterated game is more complicated, and in its complication lies hope.

The iterated game is simply the ordinary game repeated an indefinite number of times with the same players. Once again you and I face each other, with a banker sitting between. Once again we each have a hand of just two cards, labelled COOPERATE and DEFECT. Once again we move by each playing one or other of these cards and the banker shells out, or levies fines, according to the rules given above. But now, instead of that being the end of the game, we pick up our cards and prepare for another round. The successive rounds of the game give us the opportunity to build up trust or mistrust, to reciprocate or placate, forgive or avenge. In an indefinitely long game, the important point is that we can both win at the expense of the banker, rather than at the expense of one another.

After ten rounds of the game, I could theoretically have won as much as $5,000, but only if you have been extraordinarily silly (or saintly) and played COOPERATE every time, in spite of the fact that I was consistently defecting. More realistically, it is easy for each of us to pick up $3,000 of the banker's money by both playing COOPERATE on all ten rounds of the game. For this we don't have to be particularly saintly, because we can both see, from the other's past moves, that the other is to be trusted. We can, in effect, police each {207} other's behaviour. Another thing that is quite likely to happen is that neither of us trusts the other: we both play DEFECT for all ten rounds of the game, and the banker gains $100 in fines from each of us. Most likely of all is that we partially trust one another, and each play some mixed sequence of COOPERATE and DEFECT, ending up with some intermediate sum of money.

The birds in Chapter 10 who removed ticks from each other's feathers were playing an iterated Prisoner's Dilemma game. How is this so? It is important, you remember, for a bird to pull off his own ticks, but he cannot reach the top of his own head and needs a companion to do that for him. It would seem only fair that he should return the favour later. But this service costs a bird time and energy, albeit not much. If a bird can get away with cheating — with having his own ticks removed but then refusing to reciprocate — he gains all the benefits without paying the costs. Rank the outcomes, and you'll find that indeed we have a true game of Prisoner's Dilemma. Both cooperating (pulling each other's ticks off) is pretty good, but there is still a temptation to do even better by refusing to pay the costs of reciprocating. Both defecting (refusing to pull ticks off) is pretty bad, but not so bad as putting effort into pulling another's ticks off and still ending up infested with ticks oneself. The payoff matrix is Figure B.

	What you do	
What I do	**Cooperate**	**Defect**
Cooperate	Fairly good **REWARD** I get my ticks removed, although I also pay the costs of removing yours	Very bad **SUCKER'S PAYOFF** I keep my ticks, while also paying the costs of removing yours
Defect	Very good **TEMPTATION**	Fairly bad **PUNISHMENT**

I get my ticks removed, and I don't pay the costs of removing yours	I keep my ticks with the small consolation of not removing yours

Figure *B*. The bird tick-removing game: payoffs to me from various outcomes {208}

But this is only one example. The more you think about it, the more you realize that life is riddled with Iterated Prisoner's Dilemma games, not just human life but animal and plant life too. Plant life? Yes, why not? Remember that we are not talking about conscious strategies (though at times we might be), but about strategies in the ‘Maynard Smithian’ sense, strategies of the kind that genes might preprogram. Later we shall meet plants, various animals and even bacteria, all playing the game of Iterated Prisoner's Dilemma. Meanwhile, let's explore more fully what is so important about iteration.

Unlike the simple game, which is rather predictable in that DEFECT is the only rational strategy, the iterated version offers plenty of strategic scope. In the simple game there are only two possible strategies, COOPERATE and DEFECT. Iteration, however, allows lots of conceivable strategies, and it is by no means obvious which one is best. The following, for instance, is just one among thousands: ‘cooperate most of the time, but on a random 10 per cent of rounds throw in a defect’. Or strategies might be conditional upon the past history of the game. My ‘Grudger’ is an example of this; it has a good memory for faces, and although fundamentally cooperative it defects if the other player has ever defected before. Other strategies might be more forgiving and have shorter memories.

Clearly the strategies available in the iterated game are limited only by our ingenuity. Can we work out which is best? This was the task that Axelrod set himself. He had the entertaining idea of running a competition, and he advertised for experts in games theory to submit strategies. Strategies, in this sense, are preprogrammed rules for action, so it was appropriate for contestants to send in their entries in computer language. Fourteen strategies were submitted. For good measure Axelrod added a fifteenth, called Random, which simply played COOPERATE and DEFECT randomly, and served as a kind of baseline ‘non-strategy’, if a strategy can't do better than Random, it must be pretty bad.

Axelrod translated all 15 strategies into one common programming language, and set them against one another in one big computer. Each strategy was paired off in turn with every other one (including a copy of itself) to play Iterated Prisoner's Dilemma. Since there were 15 strategies, there were 15 × 15, or 225 separate games going on in the computer. When each pairing had gone {209} through 200 moves of the game, the winnings were totalled up and the winner declared.

We are not concerned with which strategy won against any particular opponent. What matters is which strategy accumulated the most ‘money’, summed over all its 15 pairings. ‘Money’ means simply ‘points’, awarded according to the following scheme: mutual Cooperation, 3 points; Temptation to defect, 5 points; Punishment for mutual defection, 1 point (equivalent to a light fine in our earlier game); Sucker's payoff, 0 points (equivalent to a heavy fine in our earlier game).

		What you do	
		Cooperate	**Defect**
What I do	**Cooperate**	Fairly good **REWARD**	Very bad **SUCKER'S PAYOFF**

	for mutual cooperation 3 points	0 points
Defect	Very good **TEMPTATION** to defect 3 points	Fairly bad **PUNISHMENT** for mutual defection 1 point

Figure C. Axelrod's computer tournament: payoffs to me from various outcomes

The maximum possible score that any strategy could achieve was 15,000 (200 rounds at 5 points per round, for each of 15 opponents). The minimum possible score was 0. Needless to say, neither of these two extremes was realized. The most that a strategy can realistically hope to win in an average one of its 15 pairings cannot be much more than 600 points. This is what two players would each receive if they both consistently cooperated, scoring 3 points for each of the 200 rounds of the game. If one of them succumbed to the temptation to defect, it would very probably end up with fewer points than 600 because of retaliation by the other player (most of the submitted strategies had some kind of retaliatory behaviour built into them). We can use 600 as a kind of benchmark for a game, and express all {210} scores as a percentage of this benchmark. On this scale it is theoretically possible to score up to 166 per cent (1,000 points), but in practice no strategy's average score exceeded 600.

Remember that the ‘players’ in the tournament were not humans but computer programs, preprogrammed strategies. Their human authors played the same role as genes programming bodies (think of Chapter 4's computer chess and the Andromeda computer). You can think of the strategies as miniature ‘proxies’ for their authors. Indeed, one author could have submitted more than one strategy (although it would have been cheating — and Axelrod would presumably not have allowed it — for an author to ‘pack’ the competition with strategies, one of which received the benefits of sacrificial cooperation from the others).

Some ingenious strategies were submitted, though they were, of course, far less ingenious than their authors. The winning strategy, remarkably, was the simplest and superficially least ingenious of all. It was called Tit for Tat, and was submitted by Professor Anatol Rapoport, a well-known psychologist and games theorist from Toronto. Tit for Tat begins by cooperating on the first move and thereafter simply copies the previous move of the other player.

How might a game involving Tit for Tat proceed? As ever, what happens depends upon the other player. Suppose, first, that the other player is also Tit for Tat (remember that each strategy played against copies of itself as well as against the other 14). Both Tit for Tats begin by cooperating. In the next move, each player copies the other's previous move, which was COOPERATE. Both continue to COOPERATE until the end of the game, and both end up with the full 100 per cent ‘benchmark’ score of 600 points.

Now suppose Tit for Tat plays against a strategy called Naive Prober. Naive Prober wasn't actually entered in Axelrod's competition, but it is instructive nevertheless. It is basically identical to Tit for Tat except that, once in a while, say on a random one in ten moves, it throws in a gratuitous defection and claims the high Temptation score. Until Naive Prober tries one of its probing defections the players might as well be two Tit for Tats. A long and mutually profitable sequence of cooperation seems set to run its course, with a comfortable 100 per cent benchmark score for both players. But suddenly, without warning, say on the eighth move, Naive Prober defects. Tit for Tat, of course, has played COOPERATE on this move, and so is landed with the Sucker's payoff of 0 points. {211} Naive Prober appears to have done well, since it has obtained 5 points from that move. But in the next move Tit for Tat ‘retaliates’. It plays DEFECT, simply following its rule of imitating the opponent's previous move. Naive Prober meanwhile, blindly following its own built-in copying rule, has copied its opponent's

COOPERATE move. So it now collects the Sucker's payoff of 0 points, while Tit for Tat gets the high score of 5. In the next move, Naive Prober — rather unjustly one might think — ‘retaliates’ against Tit for Tat's defection. And so the alternation continues. During these alternating runs both players receive on average 2.5 points per move (the average of 5 and 0). This is lower than the steady 3 points per move that both players can amass in a run of mutual cooperation (and, by the way, this is the reason for the ‘additional condition’ left unexplained on page 204). So, when Naive Prober plays against Tit for Tat, both do worse than when Tit for Tat plays against another Tit for Tat. And when Naive Prober plays against another Naive Prober, both tend to do, if anything, even worse still, since runs of reverberating defection tend to get started earlier.

Now consider another strategy, called Remorseful Prober. Remorseful Prober is like Naive Prober, except that it takes active steps to break out of runs of alternating recrimination. To do this it needs a slightly longer ‘memory’ than either Tit for Tat or Naive Prober. Remorseful Prober remembers whether it has just spontaneously defected, and whether the result was prompt retaliation. If so, it ‘remorsefully’ allows its opponent ‘one free hit’ without retaliating. This means that runs of mutual recrimination are nipped in the bud. If you now work through an imaginary game between Remorseful Prober and Tit for Tat, you'll find that runs of would-be mutual retaliation are promptly scotched. Most of the game is spent in mutual cooperation, with both players enjoying the consequent generous score. Remorseful Prober does better against Tit for Tat than Naive Prober does, though not as well as Tit for Tat does against itself.

Some of the strategies entered in Axelrod's tournament were much more sophisticated than either Remorseful Prober or Naive Prober, but they too ended up with fewer points, on average, than simple Tit for Tat. Indeed the least successful of all the strategies (except Random) was the most elaborate. It was submitted by ‘Name withheld’ — a spur to pleasing speculation: Some *eminence grise* in the Pentagon? The head of the CIA? Henry Kissinger? Axelrod himself? I suppose we shall never know. {212}

It isn't all that interesting to examine the details of the particular strategies that were submitted. This isn't a book about the ingenuity of computer programmers. It is more interesting to classify strategies according to certain categories, and examine the success of these broader divisions. The most important category that Axelrod recognizes is ‘nice’. A nice strategy is defined as one that is never the first to defect. Tit for Tat is an example. It is capable of defecting, but it does so only in retaliation. Both Naive Prober and Remorseful Prober are nasty strategies because they sometimes defect, however rarely, when not provoked. Of the 15 strategies entered in the tournament, 8 were nice. Significantly, the 8 top-scoring strategies were the very same 8 nice strategies, the 7 nasties trailing well behind. Tit for Tat obtained an average of 504.5 points: 84 per cent of our benchmark of 600, and a good score. The other nice strategies scored only slightly less, with scores ranging from 83.4 per cent down to 786 per cent. There is a big gap between this score and the 66.8 per cent obtained by Graaskamp, the most successful of all the nasty strategies. It seems pretty convincing that nice guys do well in this game.

Another of Axelrod's technical terms is ‘forgiving’. A forgiving strategy is one that, although it may retaliate, has a short memory. It is swift to overlook old misdeeds. Tit for Tat is a forgiving strategy. It raps a defector over the knuckles instantly but, after that, lets bygones be bygones. Chapter 10's Grudger is totally unforgiving. Its memory lasts the entire game. It never forgets a grudge against a player who has ever defected against it, even once. A strategy formally identical to Grudger was entered in Axelrod's tournament under the name of Friedman, and it didn't do particularly well. Of all the nice strategies (note that it is technically nice, although it is totally unforgiving), Grudger/Friedman did next to worst. The reason unforgiving strategies don't do very well is that they can't break out of runs of mutual recrimination, even when their opponent is ‘remorseful’.

It is possible to be even more forgiving than Tit for Tat. Tit for Two Tats allows its opponents two defections in a row before it eventually retaliates. This might seem excessively saintly and magnanimous. Nevertheless Axelrod worked out that, if only somebody had submitted Tit for Two Tats, it would have won the tournament. This is because it is so good at avoiding runs of mutual recrimination. {213}

So, we have identified two characteristics of winning strategies: niceness and forgivingness. This almost utopian-sounding conclusion — that niceness and forgivingness pay — came as a surprise to many of the experts, who had tried to be too cunning by submitting subtly nasty strategies; while even those who had submitted nice strategies had not dared anything so forgiving as Tit for Two Tats.

Axelrod announced a second tournament. He received 62 entries and again added Random, making 63 in all. This time, the exact number of moves per game was not fixed at 200 but was left open, for a good reason that I shall come to later. We can still express scores as a percentage of the 'benchmark', or 'always cooperate' score, even though that benchmark needs more complicated calculation and is no longer a fixed 600 points.

Programmers in the second tournament had all been provided with the results of the first, including Axelrod's analysis of why Tit for Tat and other nice and forgiving strategies had done so well. It was only to be expected that the contestants would take note of this background information, in one way or another. In fact, they split into two schools of thought. Some reasoned that niceness and forgivingness were evidently winning qualities, and they accordingly submitted nice, forgiving strategies. John Maynard Smith went so far as to submit the super-forgiving Tit for Two Tats. The other school of thought reasoned that lots of their colleagues, having read Axelrod's analysis, would now submit nice, forgiving strategies. They therefore submitted nasty strategies, trying to exploit these anticipated softies!

But once again nastiness didn't pay. Once again, Tit for Tat, submitted by Anatol Rapoport, was the winner, and it scored a massive 96 per cent of the benchmark score. And again nice strategies, in general, did better than nasty ones. All but one of the top 15 strategies were nice, and all but one of the bottom 15 were nasty. But although the saintly Tit for Two Tats would have won the first tournament if it had been submitted, it did not win the second. This was because the field now included more subtle nasty strategies capable of preying ruthlessly upon such an out-and-out softy.

This underlines an important point about these tournaments. Success for a strategy depends upon which other strategies happen to be submitted. This is the only way to account for the difference between the second tournament, in which Tit for Two Tats was {214} ranked well down the list, and the first tournament, which Tit for Two Tats would have won. But, as I said before, this is not a book about the ingenuity of computer programmers. Is there an objective way in which we can judge which is the truly best strategy, in a more general and less arbitrary sense? Readers of earlier chapters will already be prepared to find the answer in the theory of evolutionarily stable strategies.

I was one of those to whom Axelrod circulated his early results, with an invitation to submit a strategy for the second tournament. I didn't do so, but I did make another suggestion. Axelrod had already begun to think in ESS terms, but I felt that this tendency was so important that I wrote to him suggesting that he should get in touch with W. D. Hamilton, who was then, though Axelrod didn't know it, in a different department of the same university, the University of Michigan. He did indeed immediately contact Hamilton, and the result of their subsequent collaboration was a brilliant joint paper published in the journal *Science* in 1981, a paper that won the Newcomb Cleveland Prize of the American Association for the Advancement of Science. In addition to discussing some delightfully way-out biological examples of iterated prisoner's dilemmas, Axelrod and Hamilton gave what I regard as due recognition to the ESS approach.

Contrast the ESS approach with the ‘round-robin’ system that Axelrod's two tournaments followed. A round-robin is like a football league. Each strategy was matched against each other strategy an equal number of times. The final score of a strategy was the sum of the points it gained against all the other strategies. To be successful in a round-robin tournament, therefore, a strategy has to be a good competitor against all the other strategies that people *happen* to have submitted. Axelrod's name for a strategy that is good against a wide variety of other strategies is ‘robust’. Tit for Tat turned out to be a robust strategy. But the set of strategies that people happen to have submitted is an arbitrary set. This was the point that worried us above. It just so happened that in Axelrod's original tournament about half the entries were nice. Tit for Tat won in this climate, and Tit for Two Tats would have won in this climate if it had been submitted. But suppose that nearly all the entries had just happened to be nasty. This could very easily have occurred. After all, 6 out of the 14 strategies submitted were nasty. If 13 of them had been nasty, Tit for Tat wouldn't have won. The ‘climate’ would have been wrong {215} for it. Not only the money won, but the rank order of success among strategies, depends upon which strategies happen to have been submitted; depends, in other words, upon something as arbitrary as human whim. How can we reduce this arbitrariness? By ‘thinking ESS’.

The important characteristic of an evolutionarily stable strategy, you will remember from earlier chapters, is that it carries on doing well when it is already numerous in the population of strategies. To say that Tit for Tat, say, is an ESS, would be to say that Tit for Tat does well in a climate dominated by Tit for Tat. This could be seen as a special kind of ‘robustness’. As evolutionists we are tempted to see it as the only kind of robustness that matters. Why does it matter so much? Because, in the world of Darwinism, winnings are not paid out as money; they are paid out as offspring. To a Darwinian, a successful strategy is one that has become numerous in the population of strategies. For a strategy to remain successful, it must do well specifically when it is numerous, that is in a climate dominated by copies of itself.

Axelrod did, as a matter of fact, run a third round of his tournament as natural selection might have run it, looking for an ESS. Actually he didn't call it a third round, since he didn't solicit new entries but used the same 63 as for Round 2. I find it convenient to treat it as Round 3, because I think it differs from the two ‘round-robin’ tournaments more fundamentally than the two round-robin tournaments differ from each other.

Axelrod took the 63 strategies and threw them again into the computer to make ‘generation 1’ of an evolutionary succession. In ‘generation 1’, therefore, the ‘climate’ consisted of an equal representation of all 63 strategies. At the end of generation 1, winnings to each strategy were paid out, not as ‘money’ or ‘points’, but as *offspring*, identical to their (asexual) parents. As generations went by, some strategies became scarcer and eventually went extinct. Other strategies became more numerous. As the proportions changed, so, consequently, did the ‘climate’ in which future moves of the game took place.

Eventually, after about 1,000 generations, there were no further changes in proportions, no further changes in climate. Stability was reached. Before this, the fortunes of the various strategies rose and fell, just as in my computer simulation of the Cheats, Suckers, and Grudgers. Some of the strategies started going extinct from the start, {216} and most were extinct by generation 200. Of the nasty strategies, one or two of them began by increasing in frequency, but their prosperity, like that of Cheat in my simulation, was short-lived. The only nasty strategy to survive beyond generation 200 was one called Harrington. Harrington's fortunes rose steeply for about the first 150 generations. Thereafter it declined rather gradually, approaching extinction around generation 1,000. Harrington did well temporarily for the same reason as my original Cheat did. It exploited softies like Tit for Two Tats (too forgiving) while these were still around. Then, as the softies were driven extinct, Harrington followed them, having no easy prey left. The field was free for ‘nice’ but ‘provocable’ strategies like Tit for Tat.

Tit for Tat itself, indeed, came out top in five out of six runs of Round 3, just as it had in Rounds 1 and 2. Five other nice but provocable strategies ended up nearly as successful (frequent in the population) as Tit for Tat; indeed, one of them won the sixth run. When all the nasties had been driven extinct, there was no way in which any of the nice strategies could be distinguished from Tit for Tat or from each other, because they all, being nice, simply played COOPERATE against each other.

A consequence of this indistinguishability is that, although Tit for Tat seems like an ESS, it is strictly not a true ESS. To be an ESS, remember, a strategy must not be invadable, when it is common, by a rare, mutant strategy. Now it is true that Tit for Tat cannot be invaded by any nasty strategy, but another nice strategy is a different matter. As we have just seen, in a population of nice strategies they will all look and behave exactly like one another: they will all COOPERATE all the time. So any other nice strategy, like the totally saintly Always Cooperate, although admittedly it will not enjoy a positive selective advantage over Tit for Tat, can nevertheless drift into the population without being noticed. So technically Tit for Tat is not an ESS.

You might think that since the world stays just as nice, we could as well regard Tit for Tat as an ESS. But alas, look what happens next. Unlike Tit for Tat, Always Cooperate is not stable against invasion by nasty strategies such as Always Defect. Always Defect does well against Always Cooperate, since it gets the high ‘Temptation’ score every time. Nasty strategies like Always Defect will come in to keep down the numbers of too nice strategies like Always Cooperate.

But although Tit for Tat is strictly speaking not a true ESS, it is {217} probably fair to treat some sort of mixture of basically nice but retaliatory ‘Tit for Tat-like’ strategies as roughly equivalent to an ESS in practice. Such a mixture might include a small admixture of nastiness. Robert Boyd and Jeffrey Lorberbaum, in one of the more interesting follow-ups to Axelrod's work, looked at a mixture of Tit for Two Tats and a strategy called Suspicious Tit for Tat. Suspicious Tit for Tat is technically nasty, but it is not *very* nasty. It behaves just like Tit for Tat itself after the first move, but — this is what makes it technically nasty — it does defect on the very first move of the game. In a climate entirely dominated by Tit for Tat, Suspicious Tit for Tat does not prosper, because its initial defection triggers an unbroken run of mutual recrimination. When it meets a Tit for Two Tats player, on the other hand, Tit for Two Tats's greater forgivingness nips this recrimination in the bud. Both players end the game with at least the ‘benchmark’, all C, score and with Suspicious Tit for Tat scoring a bonus for its initial defection. Boyd and Lorberbaum showed that a population of Tit for Tat could be invaded, evolutionarily speaking, by a *mixture* of Tit for Two Tats and Suspicious Tit for Tat, the two prospering in each other's company. This combination is almost certainly not the only combination that could invade in this kind of way. There are probably lots of mixtures of slightly nasty strategies with nice and very forgiving strategies that are together capable of invading. Some might see this as a mirror for familiar aspects of human life.

Axelrod recognized that Tit for Tat is not strictly an ESS, and he therefore coined the phrase ‘collectively stable strategy’ to describe it. As in the case of true ESSs, it is possible for more than one strategy to be collectively stable at the same time. And again, it is a matter of luck which one comes to dominate a population. Always Defect is also stable, as well as Tit for Tat.

In a population that has already come to be dominated by Always Defect, no other strategy does better. We can treat the system as bistable, with Always Defect being one of the stable points, Tit for Tat (or some mixture of mostly nice, retaliatory strategies) the other stable point. Whichever stable point comes to dominate the population first will tend to stay dominant.

But what does 'dominate' mean, in quantitative terms? How many Tit for Tats must there be in order for Tit for Tat to do better than Always Defect? That depends upon the detailed payoffs that the banker has agreed to shell out in this particular game. All we can say {218} in general is that there is a critical frequency, a knife-edge. On one side of the knife-edge the critical frequency of Tit for Tat is exceeded, and selection will favour more and more Tit for Tats. On the other side of the knife-edge the critical frequency of Always Defect is exceeded, and selection will favour more and more Always Defects. We met the equivalent of this knife-edge, you will remember, in the story of the Grudgers and Cheats in Chapter 10.

It obviously matters, therefore, on which side of the knife-edge a population happens to *start*. And we need to know how it might happen that a population could occasionally cross from one side of the knife-edge to the other. Suppose we start with a population already sitting on the Always Defect side. The few Tit for Tat individuals don't meet each other often enough to be of mutual benefit. So natural selection pushes the population even further towards the Always Defect extreme. If only the population could just manage, by random drift, to get itself over the knife-edge, it could coast down the slope to the Tit for Tat side, and everyone would do much better at the banker's (or 'nature's') expense. But of course populations have no group will, no group intention or purpose. They cannot strive to leap the knife-edge. They will cross it only if the undirected forces of nature happen to lead them across.

How could this happen? One way to express the answer is that it might happen by 'chance'. But 'chance' is just a word expressing ignorance. It means 'determined by some as yet unknown, or unspecified, means'. We can do a little better than 'chance'. We can try to think of practical ways in which a minority of Tit for Tat individuals might happen to increase to the critical mass. This amounts to a quest for possible ways in which Tit for Tat individuals might happen to cluster together in sufficient numbers that they can all benefit at the banker's expense.

This line of thought seems to be promising, but it is rather vague. How exactly might mutually resembling individuals find themselves clustered together, in local aggregations? In nature, the obvious way is through genetic relatedness — kinship. Animals of most species are likely to find themselves living close to their sisters, brothers and cousins, rather than to random members of the population. This is not necessarily through choice. It follows automatically from ‘viscosity’ in the population. Viscosity means any tendency for individuals to continue living close to the place where they were born. For instance, through most of history, and in most parts of the {219} world (though not, as it happens, in our modern world), individual humans have seldom strayed more than a few miles from their birthplace. As a result, local clusters of genetic relatives tend to build up. I remember visiting a remote island off the west coast of Ireland, and being struck by the fact that almost everyone on the island had the most enormous jug-handle ears. This could hardly have been because large ears suited the climate (there are strong offshore winds). It was because most of the inhabitants of the island were close kin of one another.

Genetic relatives will tend to be alike not just in facial features but in all sorts of other respects as well. For instance, they will tend to resemble each other with respect to genetic tendencies to play — or not to play — Tit for Tat. So even if Tit for Tat is rare in the population as a whole, it may still be locally common. In a local area, Tit for Tat individuals may meet each other often enough to prosper from mutual cooperation, even though calculations that take into account only the global frequency in the total population might suggest that they are below the ‘knife-edge’ critical frequency.

If this happens, Tit for Tat individuals, cooperating with one another in cosy little local enclaves, may prosper so well that they grow from small local clusters into larger local clusters. These local clusters may grow so large that they spread out into other areas, areas that had hitherto been dominated, numerically, by individuals playing Always Defect. In thinking of these local enclaves, my Irish island is a misleading parallel because it is physically cut off. Think, instead, of a large population in which there is not much movement, so that individuals tend to resemble their immediate neighbours more than their more distant neighbours, even though there is continuous interbreeding all over the whole area.

Coming back to our knife-edge, then, Tit for Tat could surmount it. All that is required is a little local clustering, of a sort that will naturally tend to arise in natural populations. Tit for Tat has a built-in gift, even when rare, for crossing the knife-edge over to its own side. It is as though there were a secret passage underneath the knife-edge. But that secret passage contains a one-way valve: there is an asymmetry. Unlike Tit for Tat, Always Defect, though a true ESS, cannot use local clustering to cross the knife-edge. On the contrary. Local clusters of Always Defect individuals, far from prospering by each other's presence, do especially *badly* in each other's presence. Far from quietly helping one another at the {220} expense of the banker, they do one another down. Always Defect, then, unlike Tit for Tat, gets no help from kinship or viscosity in the population.

So, although Tit for Tat may be only dubiously an ESS, it has a sort of higher-order stability. What can this mean? Surely, stable is stable. Well, here we are taking a longer view. Always Defect resists invasion for a long time. But if we wait long enough, perhaps thousands of years, Tit for Tat will eventually muster the numbers required to tip it over the knife-edge, and the population will flip. But the reverse will not happen. Always Defect, as we have seen, cannot benefit from clustering, and so does not enjoy this higher-order stability.

Tit for Tat, as we have seen, is ‘nice’, meaning never the first to defect, and ‘forgiving’, meaning that it has a short memory for past misdeeds. I now introduce another of Axelrod's evocative technical terms. Tit for Tat is also ‘not envious’. To be *envious*, in Axelrod's terminology, means to strive for more money than the other player, rather than for an absolutely large quantity of the banker's money. To be non-envious means to be quite happy if the other player wins just as much money as you do, so long as you both thereby win more from the banker. Tit for Tat never actually ‘wins’ a game. Think about it and you'll see that it *cannot* score more than its ‘opponent’ in any particular game because it never defects except in retaliation. The most it can do is draw with its opponent. But it tends to achieve each draw with a high, shared score. Where Tit for Tat and other nice strategies are concerned, the very word ‘opponent’ is inappropriate. Sadly, however, when psychologists set up games of Iterated Prisoner's Dilemma between real humans, nearly all players succumb to envy and therefore do relatively poorly in terms of money. It seems that many people, perhaps without even thinking about it, would rather do down the other player than cooperate with the other player to do down the banker. Axelrod's work has shown what a mistake this is.

It is only a mistake in certain kinds of game. Games theorists divide games into ‘zero sum’ and ‘nonzero sum’. A zero sum game is one in which a win for one player is a loss for the other. Chess is zero sum, because the aim of each player is to win, and this means to make the other player lose. Prisoner's Dilemma, however, is a nonzero sum game. There is a banker paying out money, and it is possible for the two players to link arms and laugh all the way to the bank. {221}

This talk of laughing all the way to the bank reminds me of a delightful line from Shakespeare:

> The first thing we do, let's kill all the lawyers.
>
> *2 Henry VI*

In what are called civil ‘disputes’ there is often in fact great scope for cooperation. What looks like a zero sum confrontation can, with a little goodwill, be transformed into a mutually beneficial nonzero sum game. Consider divorce. A good marriage is obviously a nonzero sum game, brimming with mutual cooperation. But even when it breaks down there are all sorts of reasons why a couple could benefit by continuing to cooperate, and treating their divorce, too, as nonzero sum. As if child welfare were not a sufficient reason, the fees of two lawyers will make a nasty dent in the family finances. So obviously a sensible and civilized couple begin by going *together* to see one lawyer, don't they?

Well, actually no. At least in England and, until recently, in all fifty states of the USA, the law, or more strictly — and significantly — the lawyers' own professional code, doesn't allow them to. Lawyers must accept only one member of a couple as a client. The other person is turned from the door, and either has no legal advice at all or is forced to go to another lawyer. And that is when the fun begins. In separate chambers but with one voice, the two lawyers immediately start referring to 'us' and 'them'. 'Us', you understand, doesn't mean me and my wife; it means me and my lawyer against her and her lawyer. When the case comes to court, it is actually listed as 'Smith *versus* Smith'! It is *assumed* to be adversarial, whether the couple feel adversarial or not, whether or not they have specifically agreed that they want to be sensibly amicable. And who benefits from treating it as an 'I win, you lose' tussle? The chances are, only the lawyers.

The hapless couple have been dragged into a zero sum game. For the lawyers, however, the case of *Smith* v. *Smith* is a nice fat nonzero sum game, with the Smiths providing the payoffs and the two professionals milking their clients' joint account in elaborately coded cooperation. One way in which they cooperate is to make proposals that they both know the other side will not accept. This prompts a counter proposal that, again, both know is unacceptable. And so it goes on. Every letter, every telephone call exchanged between the cooperating 'adversaries' adds another wad to the bill. With luck, this procedure can be dragged out for months or even years, with costs {222} mounting in parallel. The lawyers don't get together to work all this out. On the contrary, it is ironically their scrupulous separateness that is the chief instrument of their cooperation at the expense of the clients. The lawyers may not even be aware of what they are doing. Like the vampire bats that we shall meet in a moment, they are playing to well-ritualized rules. The system works without any conscious overseeing or organizing. It is all geared to forcing us into zero sum games. Zero sum for the clients, but very much nonzero sum for the lawyers.

What is to be done? The Shakespeare option is messy. It would be cleaner to get the law changed. But most parliamentarians are drawn from the legal profession, and have a zero sum mentality. It is hard to imagine a more adversarial atmosphere than the British House of Commons. (The law courts at least preserve the decencies of debate. As well they might, since ‘my learned friend and I’ are cooperating very nicely all the way to the bank.) Perhaps well-meaning legislators and, indeed, contrite lawyers should be taught a little game theory. It is only fair to add that some lawyers play exactly the opposite role, persuading clients who are itching for a zero sum fight that they would do better to reach a nonzero sum settlement out of court.

What about other games in human life? Which are zero sum and which nonzero sum? And — because this is not the same thing — which aspects of life do we *perceive* as zero or nonzero sum? Which aspects of human life foster ‘envy’, and which foster cooperation against a ‘banker’? Think, for instance, about wage-bargaining and ‘differentials’. When we negotiate our pay-rises, are we motivated by ‘envy’, or do we cooperate to maximize our real income? Do we assume, in real life as well as in psychological experiments, that we are playing a zero sum game when we are not? I simply pose these difficult questions. To answer them would go beyond the scope of this book.

Football is a zero sum game. At least, it usually is. Occasionally it can become a nonzero sum game. This happened in 1977 in the English Football League (Association Football or ‘Soccer'; the other games called football — Rugby Football, Australian Football, American Football, Irish Football, etc., are also normally zero sum games). Teams in the Football League are split into four divisions. Clubs play against other clubs within their own division, accumulating points for each win or draw throughout the season. To be in the First Division is prestigious, and also lucrative for a club since it ensures {223} large crowds. At the end of each season, the bottom three clubs in the First Division are relegated to the Second Division for the next season. Relegation seems to be regarded as a terrible fate, worth going to great efforts to avoid.

May 18th 1977 was the last day of that year's football season. Two of the three relegations from the First Division had already been determined, but the third relegation was still in contention. It would definitely be one of three teams, Sunderland, Bristol, or Coventry. These three teams, then, had everything to play for on that Saturday. Sunderland were playing against a fourth team (whose tenure in the First Division was not in doubt). Bristol and Coventry happened to be playing against each other. It was known that, if Sunderland lost their game, then Bristol and Coventry needed only to draw against each other in order to stay in the First Division. But if Sunderland won, then the team relegated would be either Bristol or Coventry, depending on the outcome of their game against each other. The two crucial games were theoretically simultaneous. As a matter of fact, however, the Bristol-Coventry game happened to be running five minutes late. Because of this, the result of the Sunderland game became known before the end of the Bristol-Coventry game. Thereby hangs this whole complicated tale.

For most of the game between Bristol and Coventry the play was, to quote one contemporary news report, ‘fast and often furious’, an exciting (if you like that sort of thing) ding-dong battle. Some brilliant goals from both sides had seen to it that the score was 2-all by the eightieth minute of the match. Then, two minutes before the end of the game, the news came through from the other ground that Sunderland had lost. Immediately, the Coventry team manager had the news flashed up on the giant electronic message board at the end of the ground. Apparently all 22 players could read, and they all realized that they needn't bother to play hard any more. A draw was all that either team needed in order to avoid relegation. Indeed, to put effort into scoring goals was now positively bad policy since, by taking players away from defence, it carried the risk of actually losing — and being relegated after all. Both sides became intent on securing a draw. To quote the same news report: ‘Supporters who had been fierce rivals seconds before when Don Gillies fired in an 80th minute equaliser for Bristol, suddenly joined in a combined celebration. Referee Ron Challis watched helpless as the players pushed the ball around with little or no challenge to the man in {224} possession.’ What had previously been a zero sum game had suddenly, because of a piece of news from the outside world, become a nonzero sum game. In the terms of our earlier discussion, it is as if an external ‘banker’ had magically appeared, making it possible for both Bristol and Coventry to benefit from the same outcome, a draw.

Spectator sports like football are normally zero sum games for a good reason. It is more exciting for crowds to watch players striving mightily against one another than to watch them conniving amicably. But real life, both human life and plant and animal life, is not set up for the benefit of spectators. Many situations in real life are, as a matter of fact, equivalent to nonzero sum games. Nature often plays the role of ‘banker’, an4 individuals can therefore benefit from one another's success. They do not have to do down rivals in order to benefit themselves. Without departing from the fundamental laws of the selfish gene, we can see how cooperation and mutual assistance can flourish even in a basically selfish world. We can see how, in Axelrod's meaning of the term, nice guys may finish first.

But none of this works unless the game is *iterated*. The players must know (or ‘know’) that the present game is not the last one between them. In Axelrod's haunting phrase, the ‘shadow of the future’ must be long. But how long must it be? It can't be infinitely long. From a theoretical point of view it doesn't matter how long the game is; the important thing is that neither player should *know* when the game is going to end. Suppose you and I were playing against each other, and suppose we both knew that the number of rounds in the game was to be exactly 100. Now we both understand that the 100th round, being the last, will be equivalent to a simple one-off game of Prisoner's Dilemma. Therefore the only rational strategy for either of us to play on the 100th round will be DEFECT, and we can each assume that the other player will work that out and be fully resolved to defect on the last round. The last round can therefore be written off as predictable. But now the 99th round will be the equivalent of a one-off game, and the only rational choice for each player on this last but one game is also DEFECT. The 98th round succumbs to the same reasoning, and so on back. Two strictly rational players, each of whom assumes that the other is strictly rational, can do nothing but defect if they both know how many rounds the game is destined to run. For this reason, when games theorists talk about the Iterated or Repeated Prisoner's Dilemma {225} game, they always assume that the end of the game is unpredictable, or known only to the banker.

Even if the exact number of rounds in the game is not known for certain, in real life it is often possible to make a statistical guess as to how much longer the game is *likely* to last. This assessment may become an important part of strategy. If I notice the banker fidget and look at his watch, I may well conjecture that the game is about to be brought to an end, and I may therefore feel tempted to defect. If I suspect that you too have noticed the banker fidgeting, I may fear that you too may be contemplating defection. I will probably be anxious to get my defection in first. Especially since I may fear that you are fearing that I. . .

The mathematician's simple distinction between the one-off Prisoner's Dilemma game and the Iterated Prisoner's Dilemma game is too simple. Each player can be expected to behave as if he possessed a continuously updated estimate of how long the game is likely to go on. The longer his estimate, the more he will play according to the mathematician's expectations for the true iterated game: in other words, the nicer, more forgiving, less envious he will be. The shorter his estimate of the future of the game, the more he will be inclined to play according to the mathematician's expectations for the one-off game: the nastier, and less forgiving will he be.

Axelrod draws a moving illustration of the importance of the shadow of the future from a remarkable phenomenon that grew up during the First World War, the so-called live-and-let-live system. His source is the research of the historian and sociologist Tony Ashworth. It is quite well known that at Christmas British and German troops briefly fraternized and drank together in no-man's-land. Less well known, but in my opinion more interesting, is the fact that unofficial and unspoken non-aggression pacts, a ‘live-and-let-live’ system, flourished all up and down the front lines for at least two years starting in 1914. A senior British officer, on a visit to the trenches, is quoted as being astonished to observe German soldiers walking about within rifle range behind their own line. ‘Our men appeared to take no notice. I privately made up my mind to do away with that sort of thing when we took over; such things should not be allowed. These people evidently did not know there was a war on. Both sides apparently believed in the policy of "live-and-let-live".’

The theory of games and the Prisoner's Dilemma had not been invented in those days but, with hindsight, we can see pretty clearly {226} what was going on, and Axelrod provides a fascinating analysis. In the entrenched warfare of those times, the shadow of the future for each platoon was long. That is to say, each dug-in group of British soldiers could expect to be facing the same dug-in group of Germans for many months. Moreover, the ordinary soldiers never knew when, if ever, they were going to be moved; army orders are notoriously arbitrary, capricious and incomprehensible to those receiving them. The shadow of the future was quite long enough, and indeterminate enough, to foster the development of a Tit for Tat type of cooperation. Provided, that is, that the situation was equivalent to a game of Prisoner's Dilemma.

To qualify as a true Prisoner's Dilemma, remember, the payoffs have to follow a particular rank order. Both sides must see mutual cooperation (CC) as preferable to mutual defection. Defection while the other side cooperates (DC) is even better if you can get away with it. Cooperation while the other side defects (CD) is worst of all. Mutual defection (DD) is what the general staff would like to see. They want to see their own chaps, keen as mustard, potting Jerries (or Tommies) whenever the opportunity arises.

Mutual cooperation was undesirable from the generals’ point of view, because it wasn't helping them to win the war. But it was highly desirable from the point of view of the individual soldiers on both sides. They didn't want to be shot. Admittedly — and this takes care of the other payoff conditions needed to make the situation a true Prisoner's Dilemma — they probably agreed with the generals in preferring to win the war rather than lose it. But that is not the choice that faces an individual soldier. The outcome of the entire war is unlikely to be materially affected by what he, as an individual, does. Mutual cooperation with the particular enemy soldiers facing you across no-man's-land most definitely does affect your own fate, and is greatly preferable to mutual defection, even though you might, for patriotic or disciplinary reasons, marginally prefer to defect (DC) if you could get away with it. It seems that the situation was a true prisoner's dilemma. Something like Tit for Tat could be expected to grow up, and it did.

The locally stable strategy in any particular part of the trench lines was not necessarily Tit for Tat itself. Tit for Tat is one of a family of nice, retaliatory but forgiving strategies, all of which are, if not technically stable, at least difficult to invade once they arise. Three Tits for a Tat, for instance, grew up in one local area according to a contemporary account. {227}

> We go out at night in front of the trenches. . . The German working parties are also out, so it is not considered etiquette to fire. The really nasty things are rifle grenades. . . They can kill as many as eight or nine men if they do fall into a trench . . . But we never use ours unless the Germans get particularly noisy, as on their system of retaliation three for every one of ours come back.

It is important, for any member of the Tit for Tat family of strategies, that the players are punished for defection. The threat of retaliation must always be there. Displays of retaliatory capability were a notable feature of the live-and-let-live system. Crack shots on both sides would display their deadly virtuosity by firing, not at enemy soldiers, but at inanimate targets close to the enemy soldiers, a technique also used in Western films (like shooting out candle flames). It does not seem ever to have been satisfactorily answered why the two first operational atomic bombs were used — against the strongly voiced wishes of the leading physicists responsible for developing them — to destroy two cities instead of being deployed in the equivalent of spectacularly shooting out candles.

An important feature of Tit for Tat-like strategies is that they are forgiving. This, as we have seen, helps to damp down what might otherwise become long and damaging runs of mutual recrimination. The importance of damping down retaliation is dramatized by the following memoir by a British (as if the first sentence left us in any doubt) officer:

> I was having tea with A company when we heard a lot of shouting and went to investigate. We found our men and the Germans standing on their respective parapets. Suddenly a salvo arrived but did no damage. Naturally both sides got down and our men started swearing at the Germans, when all at once a brave German got on to his parapet and shouted out ‘We are very sorry about that; we hope no one was hurt. It is not our fault, it is that damned Prussian artillery.’

Axelrod comments that this apology ‘goes well beyond a merely instrumental effort to prevent retaliation. It reflects moral regret for having violated a situation of trust, and it shows concern that someone might have been hurt.' Certainly an admirable and very brave German.

Axelrod also emphasizes the importance of predictability and ritual in maintaining a stable pattern of mutual trust. A pleasing example of this was the 'evening gun' fired by British artillery with {228} clockwork regularity at a certain part of the line. In the words of a German soldier:

> At seven it came — so regularly that you could set your watch by it. . . It always had the same objective, its range was accurate, it never varied laterally or went beyond or fell short of the mark. . . There were even some inquisitive fellows who crawled out. . . a little before seven, in order to see it burst.

The German artillery did just the same thing, as the following account from the British side shows:

> So regular were they [the Germans] in their choice of targets, times of shooting, and number of rounds fired, that. . . Colonel Jones. . . knew to a minute where the next shell would fall. His calculations were very accurate, and he was able to take what seemed to uninitiated Staff Officers big risks, knowing that the shelling would stop before he reached the place being shelled.

Axelrod remarks that such 'rituals of perfunctory and routine firing sent a double message. To the high command they conveyed aggression, but to the enemy they conveyed peace.'

The live-and-let-live system could have been worked out by verbal negotiation, by conscious strategists bargaining round a table. In fact it was not. It grew up as a series of local conventions, through people responding to one another's *behaviour;* the individual soldiers were probably hardly aware that the growing up was going on. This need not surprise us. The strategies in Axelrod's computer were definitely unconscious. It was their behaviour that defined them as nice or nasty, as forgiving or unforgiving, envious or the reverse. The programmers who designed them may have been any of these things, but that is irrelevant. A nice, forgiving, non-envious strategy could easily be programmed into a computer by a very nasty man. And vice versa. A strategy's niceness is recognized by its behaviour, not by its motives (for it has none) nor by the personality of its author (who has faded into the background by the time the program is running in the computer). A computer program can behave in a strategic manner, without being aware of its strategy or, indeed, of anything at all.

We are, of course, entirely familiar with the idea of unconscious strategists, or at least of strategists whose consciousness, if any, is irrelevant. Unconscious strategists abound in the pages of this book. Axelrod's programs are an excellent model for the way we, throughout the book, have been thinking of animals and plants, and {229} indeed of genes. So it is natural to ask whether his optimistic conclusions — about the success of non-envious, forgiving niceness — also apply in the world of nature. The answer is yes, of course they do. The only conditions are that nature should sometimes set up games of Prisoner's Dilemma, that the shadow of the future should be long, and that the games should be nonzero sum games. These conditions are certainly met, all round the living kingdoms.

Nobody would ever claim that a bacterium was a conscious strategist, yet bacterial parasites are probably engaged in ceaseless games of Prisoner's Dilemma with their hosts and there is no reason why we should not attribute Axelrodian adjectives — forgiving, non-envious, and so on — to their strategies. Axelrod and Hamilton point out that normally harmless or beneficial bacteria can turn nasty, even causing lethal sepsis, in a person who is injured. A doctor might say that the person's ‘natural resistance’ is lowered by the injury. But perhaps the real reason is to do with games of Prisoner's Dilemma. Do the bacteria, perhaps, have something to gain, but usually keep themselves in check? In the game between human and bacteria, the ‘shadow of the future’ is normally long since a typical human can be expected to live for years from any given starting-point. A seriously wounded human, on the other hand, may present a potentially much shorter shadow of the future to his bacterial guests. The ‘Temptation to defect’ correspondingly starts to look like a more attractive option than the ‘Reward for mutual cooperation’. Needless to say, there is no suggestion that the bacteria work all this out in their nasty little heads! Selection on generations of bacteria has presumably built into them an unconscious rule of thumb which works by purely biochemical means.

Plants, according to Axelrod and Hamilton, may even take revenge, again obviously unconsciously. Pig trees and fig wasps share an intimate cooperative relationship. The fig that you eat is not really a fruit. There is a tiny hole at the end, and if you go into this hole (you'd have to be as small as a fig wasp to do so, and they are minute: thankfully too small to notice when you eat a fig), you find hundreds of tiny flowers lining the walls. The fig is a dark indoor hothouse for flowers, an indoor pollination chamber. And the only agents that can do the pollinating are fig wasps. The tree, then, benefits from harbouring the wasps. But what is in it for the wasps? They lay their eggs in some of the tiny flowers, which the larvae then eat. They pollinate other flowers within the same fig. ‘Defecting’, for {230} a wasp, would mean laying eggs in too many of the flowers in a fig and pollinating too few of them. But how could a fig tree ‘retaliate’? According to Axelrod and Hamilton, ‘It turns out in many cases that if a fig wasp entering a young fig does not pollinate enough flowers for seeds and instead lays eggs in almost all, the tree cuts off the developing fig at an early stage. All progeny of the wasp then perish.’

A bizarre example of what appears to be a Tit for Tat arrangement in nature was discovered by Eric Fischer in a hermaphrodite fish, the sea bass. Unlike us, these fish don't have their sex determined at conception by their chromosomes. Instead, every individual is capable of performing both female and male functions. In any one spawning episode they shed either eggs or sperm. They form monogamous pairs and, within the pair, take turns to play the male and female roles. Now, we may surmise that any individual fish, if it could get away with it, would ‘prefer’ to play the male role all the time, because the male role is cheaper. Putting it another way, an individual that succeeded in persuading its partner to play the female most of the time would gain all the benefits of ‘her’ economic investment in eggs, while ‘he’ has resources left over to spend on other things, for instance on mating with other fish.

In fact, what Fischer observed was that the fishes operate a system of pretty strict alternation. This is just what we should expect if they are playing Tit for Tat. And it is plausible that they should, because it does appear that the game is a true Prisoner's Dilemma, albeit a somewhat complicated one. To play the COOPERATE card means to play the female role when it is your turn to do so. Attempting to play the male role when it is your turn to play the female is equivalent to playing the DEFECT card. Defection is vulnerable to retaliation: the partner can refuse to play the female role next time it is ‘her’ (his?) turn to do so, or ‘she’ can simply terminate the whole relationship. Fischer did indeed observe that pairs with an uneven sharing of sex roles tended to break up.

A question that sociologists and psychologists sometimes ask is why blood donors (in countries, such as Britain, where they are not paid) give blood. I find it hard to believe that the answer lies in reciprocity or disguised selfishness in any simple sense. It is not as though regular blood donors receive preferential treatment when they come to need a transfusion. They are not even issued with little gold stars to wear. Maybe I am naive, but I find myself tempted to see it as a genuine case of pure, disinterested altruism. Be that as it may, {231} blood-sharing in vampire bats seems to fit the Axelrod model well. We learn this from the work of G. S. Wilkinson.

Vampires, as is well known, feed on blood at night. It is not easy for them to get a meal, but if they do it is likely to be a big one. When dawn comes, some individuals will have been unlucky and return completely empty, while those individuals that have managed to find a victim are likely to have sucked a surplus of blood. On a subsequent night the luck may run the other way. So, it looks like a promising case for a bit of reciprocal altruism. Wilkinson found that those individuals who struck lucky on any one night did indeed sometimes donate blood, by regurgitation, to their less fortunate comrades. Out of 110 regurgitations that Wilkinson witnessed, 77 could easily be understood as cases of mothers feeding their children, and many other instances of blood-sharing involved other kinds of genetic relatives. There still remained, however, some examples of blood-sharing among unrelated bats, cases where the ‘blood is thicker than water’ explanation would not fit the facts. Significantly the individuals involved here tended to be frequent roostmates — they had every opportunity to interact with one another repeatedly, as is required for an Iterated Prisoner's Dilemma. But were the other requirements for a Prisoner's Dilemma met? The payoff matrix in Figure D is what we should expect if they were.

		What you do	
What I do		**Cooperate**	**Defect**
	Cooperate	Fairly good **REWARD** I get blood on my unlucky nights, which saves me from starving. I have to give blood on my lucky nights, which doesn't cost me too much	Very bad **SUCKER'S PAYOFF** I pay the cost of caving your life on my good night. But on my bad night you don't feed me and I run a real risk of starving to death
	Defect	Very good **TEMPTATION** You save my life on my poor night. But	Fairly bad **PUNISHMENT** I don't have to pay the slight costs of

then I get the added benefit of not having to pay the slight cost of feeding you on my good night	feeding you on my good nights. But I run a real risk of starving on my poor nights

Figure D. Vampire bat blood-donor scheme: payoffs to me from various outcomes. {232}

Do vampire economics really conform to this table? Wilkinson looked at the rate at which starved vampires lose weight. From this he calculated the time it would take a sated bat to starve to death, the time it would take an empty bat to starve to death, and all intermediates. This enabled him to cash out blood in the currency of hours of prolonged life. He found, not really surprisingly, that the exchange rate is different, depending upon how starved a bat is. A given amount of blood adds more hours to the life of a highly starved bat than to a less starved one. In other words, although the act of donating blood would increase the chances of the donor dying, this increase was small compared with the increase in the recipient's chances of surviving. Economically speaking, then, it seems plausible that vampire economics conform to the rules of a Prisoner's Dilemma. The blood that the donor gives up is less precious to her (social groups in vampires are female groups) than the same quantity of blood is to the recipient. On her unlucky nights she really would benefit enormously from a gift of blood. But on her lucky nights she would benefit slightly, if she could get away with it, from defecting — refusing to donate blood. ‘Getting away with it’, of course, means something only if the bats are adopting some kind of Tit for Tat strategy. So, are the other conditions for the evolution of Tit for Tat reciprocation met?

In particular, can these bats recognize one another as individuals? Wilkinson did an experiment with captive bats, proving that they can. The basic idea was to take one bat away for a night and starve it while the others were all fed. The unfortunate starved bat was then returned to the roost, and Wilkinson watched to see who, if anyone, gave it food. The experiment was repeated many times, with the bats taking turns to be the starved victim. The key point was that this population of captive bats was a mixture of two separate groups, taken from caves many miles apart. If vampires are capable of recognizing their friends, the experimentally starved bat should turn out to be fed only by those from its own original cave.

That is pretty much what happened. Thirteen cases of donation were observed. In twelve out of these thirteen, the donor bat was an ‘old friend’ of the starved victim, taken from the same cave; in only one out of the thirteen cases was the starved victim fed by a ‘new friend’, not taken from the same cave. Of course this could be a coincidence but we can calculate the odds against this. They come to less than one in 500. It is pretty safe to conclude that the bats really {233} were biased in favour of feeding old friends rather than strangers from a different cave.

Vampires are great mythmakers. To devotees of Victorian Gothic they are dark forces that terrorize by night, sapping vital fluids, sacrificing an innocent life merely to gratify a thirst. Combine this with that other Victorian myth, nature red in tooth and claw, and aren't vampires the very incarnation of deepest fears about the world of the selfish gene? As for me, I am sceptical of all myths. If we want to know where the truth lies in particular cases, we have to look. What the Darwinian corpus gives us is not detailed expectations about particular organisms. It gives us something subtler and more valuable: understanding of principle. But if we must have myths, the real facts about vampires could tell a different moral tale. To the bats themselves, not only is blood thicker than water. They rise above the bonds of kinship, forming their own lasting ties of loyal blood-brotherhood. Vampires could form the vanguard of a comfortable new myth, a myth of sharing, mutualistic cooperation. They could herald the benignant idea that, even with selfish genes at the helm, nice guys can finish first.