

PART III: NATURAL HISTORY

The next three chapters are about Darwin's theory of natural selection. Chapter 7 is about how we use natural selection to make sense of biology, and chapters 8 and 9 are about theoretical issues in the theory itself, including the role of genes in inheritance, the relationship between nature and nurture, evolution and development, and the 'units of selection' controversy. These two chapters mostly deal with issues in theoretical biology, but my aim throughout is expose the underlying philosophical problems in understanding evolutionary processes in a way that will enable me, in the final part, to apply the same explanatory framework to social history.

Chapter 7

Functions and Norms

Nothing makes sense in biology, except in the light of evolution
— Dobzhansky

For a long time the biologist treated teleology as he would a woman he could not do without, but did not care to be seen with in public. The concept of [genetic] programme has made an honest woman of teleology.
— Jacob, *The Logic of Living Systems*

In chapters 5 and 6 I argued that we can define the truth and meaning of a belief in terms of the successful actions that it plays a role in. This then left the problem of determining what actions should be considered successful, and why? Darwin offers one possible answer; and in this chapter I discuss why, and how, we should use it.

7.1 Functional Explanation and Darwinian Norms

According to Darwin, a product of natural selection is successful if it contributes to the reproduction of the organism that carries it¹. This criterion of success, or *norm*, can be applied to traits, such as hearts, or particular events, such as a single beat of a heart: if a vertebrate had no heart, or if it had a coronary attack, then it would be less likely to reproduce than one that did. We can also apply this criterion of success quite naturally to *behaviour*, and to particular actions and the internal representations that underlie them. For example, a frog is more likely to reproduce if it catches a fly than if it misses. Therefore those occasions on which the neurons in its retinotopic map fire when a fly is present can be considered successes: its proto-belief about the presence of an ‘eatable thing’ was correct. (We shall see how we can extend this criterion of correctness to the products of social evolution, including human beliefs, in the final chapter.)

Why should we use this Darwinian norm rather than any other? The reason is that it makes possible a new way of making sense of systems that evolve over time. Remember that the original problem Darwin pondered was that the natural world seemed so *successful*. Every organism

¹Or, more accurately, if it contributes to the ability of that organism to pass on its genes. The issue of the relationship between the replication of genes and the reproduction of organisms will be the subject of chapters 8 and 9.

seemed perfectly adapted to its ecological niche, and every part of every organism seemed optimally designed to allow the whole to survive and flourish. Every part of nature seemed to fulfill a *function* for the whole. Paley drew the obvious conclusion that this design was God-given, but Darwin dared to suggest that this order could be explained by the mechanism of natural selection. In other words Darwin *naturalised* functional descriptions of nature (in the sense defined in chapter 3). Paley knew *that* the natural world was functionally organised, but only Darwin could explain *why*.

Naturalisation normally works ‘from below’ by showing how the behaviour of a whole is produced by the workings of its parts. But once the mechanism of natural selection is in place then a new kind of explanation is possible, namely *functional explanation*. It is now possible to explain the workings of the parts in terms of the contribution that they make to the whole. For example, why do hearts pump blood? The usual naturalised, i.e. *mechanical*, explanation is that hearts pump because they have chambers and valves and rhythmic control systems. But the functional explanation is that hearts pump because this helps the animal survive. Functional explanation is explanation from above, rather than from below; it accounts for the existence of an entity in terms of its effects, rather than its causes. This type of explanation only works because natural selection can explain why effects of that type in the past have led to the existence of entities of that type today. Hearts exist *because* they (or rather, their ancestors) pump blood, and so on. Millikan defines this as the Direct Proper Function of an entity (1984): it is what the entity exists in order to do; it is the Purpose of the entity; what it exists *for*.

The advantage of using the Darwinian norm, rather than any other, is that it enables us make sense of the history of the system. This is because we are using the same criterion of success that evolution uses. The Darwinian norm is a force that operates within nature, rather than a criterion that we are imposing on nature from without. The Darwinian norm is a *naturalised* norm. Consider this example. It is fairly crass, but it makes the distinction clear. Suppose we are contemplating a bluebell wood in spring, and choose to analyse it using the norm of ‘looking pretty’. The wood succeeds in meeting that norm, and this in turn defines functions for its various parts: the bluebells have the function of providing a splash of colour, the leaves on the trees fulfill the function of providing dappled sunlight, and the worms in the soil fulfill the function of aerating the soil and keeping the whole thing healthy. This norm tells us *something* about the wood; but it obviously tells us little about its history, about *why* it exists in its current form. For this we have to use the Darwinian norm. According to Darwin the function of the bluebells is to attract pollinating insects, the function of the leaves is to photosynthesise and shade out competitors, and the function of the worm is to make more worms: the Darwinian function explains why each of these entities exist.

The theory of natural selection allows us to use the function of an entity to explain its history. However, if we want to understand the history of a system there are two distinct questions we can ask. The first *how did the system get like it is now?*, i.e. we can ask about the past history of the system. But we can also ask *what will happen to the system in the future?*, i.e. we can ask about its *future* history. Now as long as there are no major disruptions to this system, and as long as all change is incremental and gradual, then the answer to the first question will also answer the second. Bluebells exist now because they attracted insects in the past, and in an undisturbed wood this situation is unlikely to change radically. But suppose that the bluebells were stolen by an avid

gardener, transplanted to a greenhouse, and artificially pollinated in order to grow new commercial varieties. Now what matters for the future history of the flowers is their ability to attract customers, rather than insects. The functions of an evolved system change as soon as a new selective regime comes into place. The Proper function — i.e. the function defined by past history — has remained the same, but the current Darwinian function has changed. We are using the same norm in each case — i.e. reproductive success — but in one case to understand how the system got to how it is now, and in the other to explain how it will change in the future. The same norm yields different functions in each case.

7.2 The Function of ‘Function’

What is the function of the concept of ‘function’? What explanatory purpose do we want to use it for? Is it supposed to explain the past history of a system, or its future? Most of the discussion in the philosophical literature has tried to define a single, unitary, concept of function (see for example Wright (1973), Cummins (1975), Neander (1991), Bigelow and Pargetter (1987), and Millikan (1984)(1993)(1999).) But there is no reason to expect that a single concept can answer both questions.

According to Millikan (and Wright) the function of an entity *must* account for its occurrence: it is the ability of the ancestors of an entity to fulfill that function that explains why it exists in that form today. Thus Millikan’s Proper Function is very similar to what we usually describe as the ‘purpose’ of an entity; i.e. what that entity exists *for*. In other words it is a *teleological* concept of function, similar to Aristotle’s Final Cause which ‘pulled’ natural events towards some future goal. But Darwin showed how the same functional organisation could be explained by the force of natural selection ‘pushing’ from the past. Therefore, as Mayr (1982) argues, Darwin’s teleology is not strictly the same as Aristotle’s but is more of an *as if* teleology — what he calls *teleonomy* — i.e. an explanation of why teleological descriptions work so well, rather than a naturalisation of final cause *per se*.²

It is often assumed that functions *must* be teleological. However there is another way of defining norms and functions based, not on the history of an entity, but on its effects. Such approaches are generally labelled *consequentialist*. Anscombe (1958) originally invoked this term to describe theories of ethical norms, such as Mill and Bentham’s Utilitarianism, or Aristotelian and Marxist ethics (Miller, 1981). To put it crudely, according to these theories an act is considered (ethically) correct if it contributes to a state of maximal happiness, *eudaimonia*, or communism, respectively. Conversely, if the act leads to a state that one considers bad (such as maximal unhappiness, dystopia, or to the continuance of capitalism) then the act was wrong. And if the act makes no positive contribution either way then it is ethically neutral. Given these norms for judging an act then, in certain circumstance, one can derive a non-teleological sense of ought. For example if one is given a number of choices then we can use consequentialist criteria for judging which is the best. That act is then what one *ought* to do. Thus consequentialist theories start from norms and then derive oughts, rather than *vice versa* as teleological theories do.

²Actually this is not quite true. Aristotle himself considered a proto-Darwinian theory of evolutionary teleology that had been proposed by Empedocles (see Aristotle’s *Physics*, II.8). Aristotle agrees that this would be an example of teleological causation but rejects it on purely empirical grounds since he saw no evidence of descent with modification in nature. Thus purposiveness is not an *a priori* part of Aristotle’s teleology, just an *a posteriori* explanation of it.

(Obviously the possibility of deriving oughts from norms depends on assumptions about agency, i.e. about the metaphysics of counterfactual assertions that an agent could have acted differently. Does it make sense to say that we ought to do something we are physically incapable of? This problem becomes relevant in section 11.3 where I discuss the role of utopias in forming ethical judgements, but does not affect the current discussion.)

The norms that consequentialism defines for an entity are derived from the norms of its consequences. Bentham and Mill believed that maximising happiness was a good thing so they concluded that acts that contribute to it are also good. Similarly it is the goodness of Aristotle's *eudaimonia* and of Marx's communism that enabled them to define norms for social acts. In short, ends justify means. The same basic strategy works for biological functions: we can define the correct function of an entity on the basis of its Darwinian consequences, rather than its Darwinian history. The historical function of a transplanted bluebell is to attract insects, but its consequentialist function is to attract customers.

Consequentialist and historical norms serve different purposes. They are complementary and so we should use them both. However it is interesting to note that in many cases the consequentialist approach is of more practical use to biologists than the teleological.

For example, what is the function of a behaviour that an animal learns in order to survive in novel circumstances that were not part of its evolutionary history? Millikan argues that all such lifetime innovations must be the products of innate learning mechanisms — such as those discovered underlying Skinnerian operant conditioning in *Aplysia* (Carew et al., 1983)(Hawkins et al., 1983). These innate mechanisms will have some well-defined evolutionary history, and hence a 'direct' proper function, from which the function of the lifetime innovations may be derived. Therefore the existence of derived proper functions for learned behaviours depends on there being an innate 'end to flexibility' in the brain, as Millikan puts it (1984, p46–48).

An historical (i.e. teleological) analysis of the functional organisation of brains encourages the view that the brain is divided into discrete and innate modules with highly specialised functions (Millikan, 1993, p49) — caricatured as the 'Swiss army knife' model. Now there is certainly a growing body of evidence for such modularisation (Pinker, 1998). But the plain fact is that we should not *assume* it is universal, nor are we ever to likely to be able to determine the evolutionary history of particular brain modules in practice (Lewontin, 1998). And nor must development 'bottom out' in innate mechanisms in quite the way that Millikan requires (Elman, 1995). A consequentialist view of function offers a more practical alternative. The progress that we have so far made in investigating the functional organisation of brains has mostly come from an understanding of the role that they play in the current behaviour and development of the organism, rather than their individual evolutionary history (Kennedy et al., 1995). It is still necessary to use the Darwinian norm to make sense of the functional organisation of brains, but we usually do this without knowing much about their evolutionary history. Of course once we uncover some aspect of how brains work — i.e. how they enable the animal survive and flourish and so meet the Darwinian norm — then the next obvious question is to ask how evolution produced those mechanisms. But, in purely practical terms, a consequentialist analysis of function is of more immediate use to a neuroscientist than an historical one.

7.3 The Function of Behaviour

The consequentialist view of function is also consistent with that which underpins the investigation of animal behaviour. Tinbergen (1951) famously defined four questions that the ethologist — or any other biologist — can ask of a trait: how is it caused, how does it develop, what does it do, and how did it evolve? For example, suppose we ask ‘why does our thumb move in a different way to the other fingers?’ We might give an answer in terms of the anatomy of the hand (causal), or in terms of its embryology (developmental), or in terms of its evolutionary history (historical), or in terms of what it currently enables us to do (functional). ‘All of these answers would be correct: no one would be complete’ (Hinde, 1982, p21). Tinbergen went on to explain the functional question as follows.

In attacking our problem, let us start from observables — i.e., from behaviour. But instead of studying its causes we shall study its effects; in other words, rather than look back in time, as we do when studying causation, we investigate what happens as a consequence of the observed behaviour. I should like to stress that this *is* an investigation of cause-effect relationships, which as such requires experimental study as well as observation and speculation; it differs from the study of behaviour causation merely by the fact that the observed behaviour is the cause, and that its effects are studied; we follow events with time instead of preceding events and we determine an animal’s success. (1965, p521)

And Tinbergen explicitly contrasts this investigation with that into the *origins* of behaviour, arguing that ‘they can suggest (*though no more than that*) the selection pressures that have in the past molded the species to what it is now’ [p523]. Of course there is no rigid distinction in practice between the forces that originally moulded a behaviour and those that now maintain it. But Tinbergen is arguing that there is, at least, a clear *conceptual* distinction between historical and consequentialist function, and therefore we can determine the consequentialist function that a behaviour serves now in perfect ignorance of its evolutionary past.

For the ethologist the key problem in determining the function of a behaviour is to determine which of its many causal consequences actually help the organism reproduce — and so succeed in the Darwinian sense. And, as Hinde notes, to answer this question it is strictly necessary to assess reproductive success with and without the character (1975, p6). This is rarely possible in naturally occurring populations — where behaviours tend to be universal — and so ethologists experimentally intervene to test which factors are significant. A classic example of this is Tinbergen’s investigation of the ‘fanning’ of the male stickleback which is done mainly when there are developing eggs in the nest. He found that when fanning is prevented, or when it is allowed but the nest is covered with a watch glass, the eggs die. However they develop normally in the absence of the male provided freshly aerated water is pumped over the nest, suggesting that the function of fanning is to ventilate the eggs (Tinbergen, 1951).

Millikan uses this methodological problem as an argument against Bigelow and Pargetter’s consequentialist account of function, in which they attribute a function to a trait ‘when it confers a survival enhancing propensity on a creature that possesses it’ (1987, p192). Millikan interprets this as meaning ‘a trait that enhances or would enhance survival is one where, on average over the *actual* individuals in the species, having it would produce a more fit individual than not having it’ (1993, p39). She concludes that if the trait is universal in a population, with no current

competitors, then the concept of ‘enhanced propensity of survival’ is meaningless. So if *all* male sticklebacks fan then that fanning cannot be said to enhance their fitness, since there are no non-fanning sticklebacks to be enhanced compared to.

The counter-argument to this objection depends on the fact that fitness is operationally defined as the *expected* number of offspring produced by a class of individuals; it is therefore a *disposition* to reproduce. And as with other dispositional properties problems arise if we understand them in terms of actual outcomes. Recall Carnap’s argument that if the dispositional property of ‘being soluble’ is defined as ‘dissolving when in water’ then the claim that ‘X dissolved because it was soluble’ is tautologous (section 3.3). Similarly, if fitness is defined as the actual rate of reproduction then we face the old canard that ‘survival of the fittest’ is also tautologous. Of course the answer is that ‘solubility’ describes a property of a substance *in virtue of which* it dissolves, and similarly ‘fitness’ describes properties of organisms in virtue of which they differentially reproduce. Fitness is a measure of the *reproductive power* of an organism

It may be the case that we only discover which substances are soluble by putting them in water, but we should not confuse the way that we measure a property with the facts in virtue of which an entity holds it. Similarly, the function of the stickleback fanning is not defined by the fact that their eggs died when Tinbergen covered them with a watch glass. It is the causal process of aeration (along with its effects on the eggs) that constitutes the function of the fanning, not the differences between fanned eggs and covered ones. In short, it is possible to define the consequentialist function of a trait in terms of the role that it plays in the reproduction of the organism now, even in the absence of a range of contemporary (or historical) alternatives.

7.4 Conclusion

In this chapter I have argued that we should use the Darwinian norm to judge the success of a biological trait because this norm naturalises functional explanations of natural history, and so fits the ‘idea of the good’ for science outlined in chapter 3. But we can use this norm in two ways. The first is to explain the origin of biological traits. And the second is to explain the future fate of the organisms that possess them.

In the final chapter I will argue that it is possible to use the same criterion to define norms for *social* traits — including language, science, and ethics — but this requires that a Darwinian-type theory of natural selection can be applied to social history. This is the subject of chapter 10. But before then it is necessary to define more precisely what we mean by natural selection. And this is the subject of the next two chapters.

Chapter 8

The Role of Genes in Natural Selection

The whole purpose of our search for a ‘unit of selection’ is to discover a suitable actor to play the leading role in our metaphors of purpose. We look at an adaptation and want to say, ‘It is for the good of . . .’. Our quest is for the right way to complete that sentence.

— Dawkins, *The Extended Phenotype*

Darwin’s theory of natural selection is extraordinarily simple, but the natural world to which it applies is extraordinarily complex. Herein lies the difficulty that is the subject of the next two chapters.

The theory is that whenever individuals reproduce at different rates according to their inherited and randomly mutated properties then those individuals will evolve to become adapted to their environments. But nature does not come ready-carved into individuals. Ecosystems spread and compete with others. Species compete within ecosystems. Species are divided into sub-populations and colonies. Organisms are made of many reproducing cells. Cells contain many components that have their own replicating DNA. And each DNA string contains many replicating sequences. So which of these individuals does the theory apply to? What is the correct way to complete Dawkins’ sentence? There are two possible answers.

The first answer is that of Darwin, Wallace, Mayr, Dobzhansky, D.S. Wilson, Gould, and Sober, and was generalised and given canonical form by Lewontin (1970). Sterelny and Kitcher (1988) describe this as a ‘pluralist’ description of evolution since it makes no *a priori* claim about what those individuals are; they may be genes, organisms, cells within an multi-cellular body, groups of individual, species, or even multi-specied ecosystems (Wilson, 1976)(Goodnight, 1990). All of these units can be selected according to the traits they inherit, and so can be candidates to complete Dawkins’ sentence.

The second answer is that of Fisher, Haldane, Wright, Williams, E.O. Wilson, Hamilton, and Dawkins, who insist that the bottom line of evolution is how much it changes gene frequencies and, through these, how much it changes phenotypes. These phenotypic effects can be expressed at many different levels, and in many different ways; but it is only the effects on gene frequencies that matter. Therefore genes are the ultimate beneficiaries of adaptations.

The debate between pluralism and gene-centrism in evolution is a paradigm case of the more

general debate over levels of organisation in nature. The pluralist explains evolution in terms of selection operating at many levels independently, whilst the gene-centrist reduces evolution at higher levels to the differential replication of genes. In the next two chapters I apply the general arguments of chapters 2 and 3 to this debate. My aim is to provide a synthesis of the two traditions and to show that not only are both necessary in order to understand evolution but, strictly speaking, neither makes sense without the other.

This debate is important for two reasons. The first is the implications that it has for our understanding of biological evolution — and this will be the main focus of the next two chapters. But this debate also has implications for the way that we apply Darwin’s theory to *social* evolution. And this will be the subject of chapter 10

8.1 Evolution and Mendelian Inheritance

What are the conditions necessary for evolution by natural selection? Are proteins and nucleic acid necessary, or could evolution occur in entities made of, say, silicon polymers? Intuitively there seems to be no reason why carbon is the only possible substrate. Is metabolism necessary, or could evolution occur in entities that are more inert? Again there seems to be no obvious reason why not. Are genes necessary, or could evolution occur in entities in which there were no discrete and universal genetic code? It is here that intuitions start to differ.

According to Lewontin, a change in gene frequencies is not a necessary part of the definition of evolution by natural selection. He argues that evolution occurs whenever we have the differential reproduction of individuals according to inherited and randomly mutated traits, and this definition is silent on the nature of the mechanism underlying this process: “No particular mechanism of inheritance is specified, but only a correlation in fitness between parent and offspring” (1970, p1). Dennett describes evolution, characterised in this way, as an *algorithmic* process, and a defining characteristic of an algorithmic process is that it is ‘substrate neutral’; i.e. it is one in which

the power of the procedure is due to its *logical* structure, not the causal powers of the materials used in the instantiation, just so long as those causal powers permit the prescribed steps to be followed exactly. (1995, p51)

Now there is no *a priori* reason why evolution by natural selection cannot occur in substrates other than proteins and nucleic acid, any more than cognition requires neurons or life requires carbon chains — all these phenomena are ‘multiply realizable’ (Kim, 1992). If we use this weak sense of substrate neutrality then Dennett’s claim is definitely true. However the issue becomes more complex when we try to define exactly what constitutes ‘following the prescribed steps’ of natural selection.

Darwin was able to characterise how the force of natural selection works in a remarkably simple way: if the organisms in a population that possess one characteristic are better able to survive and reproduce than the organisms with the alternative characteristic, and if [this characteristic] is passed on from parent to offspring, then the proportion of individuals with that characteristic will increase. ... Of course, some clarification is required of what it means to say that a trait is “inherited.” But as a first approximation, the conditional claim *if* there is heritable variation in fitness, *then* there will be evolution (if nothing interferes) seems to be remarkably straightforward. (Sober, 1985, p27)

Unfortunately the theory of natural selection became a little less straightforward as soon as we ‘clarify’ what inheritance means.

The first point in ‘clarifying’ inheritance is that natural selection as an account of the origin of species depends on discrete genetic, i.e. *Mendelian*, inheritance. If the inheritance and development of traits is not discrete then the result is not the origin of species, just blending, as Darwin was fully aware. It is to Thomas Huxley’s credit that, even whilst he acted as ‘Darwin’s Bulldog’ in his defence of the fact of evolution, he remained sceptical whether natural selection was its mechanism. This was because the discrete inheritance necessary for the formation of species had not been empirically demonstrated to his satisfaction — despite Darwin spending half a life-time trying to evolve distinct species through artificial selection (Desmond & Moore, 1992, p510-11). It was largely for this reason that natural selection was not generally accepted as the dominant force in evolution until it was unified with Mendelian genetics by Fisher (1930), Haldane (1932) and Wright (1931). But Dennett (1996) points out that Mendelism was a ‘triumph for substrate neutrality’ since neither Mendel, nor the early population geneticists, had any idea what the material realisation of these units of heredity was.

Mendel saved Darwin; but what saved Mendel? Until the discovery of the structure of DNA in 1953 Mendel’s genes were *abstracta*: theoretical entities defined solely in terms of observed patterns of inherited traits. Therefore their status as legitimate scientific entities rested entirely on their empirical usefulness. However the empirical success of Mendelian genetics as an explanation of the distribution of traits depends on finding traits that are discretely encoded in genes whose effects on phenotypes and fitness are relatively robust, and so fit our simple models. Mendel was extremely fortunate in his choice of organism and trait since, in coloration in peas, he happened upon a trait that met these conditions. However most traits are not due to single genes with full penetrance and complete expressivity, and so do not always follow the clear patterns of Mendelism. If Mendel had started by enquiring into, say, the height of his peas then it is unlikely that he would have found the clear patterns that allowed him to formulate his theory, since height depends on many different genes and is very sensitive to environmental changes. Mendelian analysis can not, in general, be used to determine the genotype of an individual organism for such quantitative traits (Griffith & Suzuki, 1993, p827).

Until 1953 Mendelian genetics was not a true theory (in the sense of section 3.3), but an as-yet unfalsified hypothesis. Until 1953 Mendel’s patterns of inheritance were always at risk of being exposed as nothing more than empirically useful descriptions, like Copernicus’ epicycles. But once Mendel’s *abstracta* were transformed into the *illata* of DNA then, even if counter-examples to Mendel’s laws were discovered, his laws of inheritance (and hence also Darwin’s theory of natural selection) would not be threatened. For example, Mendel’s Second law states that, in a diploid organism with haploid gametes, genes at different loci will be transmitted independently; thus if A,a and B,b are pairs of alleles at different loci, and if an organism is heterozygous at both loci, then the probabilities that a gamete will receive any of the four possible genetic combinations AB, Ab, aB, ab , are all equal. But this is not true in general (Kitcher, 1984). For example, alleles on the same chromosome will tend to be transmitted together. Only genes on non-homologous chromosomes assort independently. But because we have some understanding of how Mendel’s genes may be arranged on a chromosome, and of the mechanism underlying meiosis, we do not

dismiss Mendel as a result.

A similar argument applies to the case of the non-inheritance of acquired characteristics — the postulate that marked modern evolutionary theory's break from Lamarck. After 1953 Crick re-christened this postulate as the Central Dogma of molecular biology, which states that information cannot pass from protein to DNA but only the reverse; i.e. changes to the phenotype incurred during development will have no effect on the DNA passed on to subsequent generations through the germ-line. However it is interesting to note that Crick called this the central *dogma* of molecular biology, rather than the central *theorem*: it was a claim about how inheritance worked, not a known fact¹. It is only over the last 40 years that have we slowly discovered the extent to which the Central Dogma is true. Indeed, there are some isolated cases in which it is *not* strictly true, but this does not threaten the Central Dogma — or Darwin. Because we now understand the biochemistry of how DNA is replicated the Central Dogma can now be regarded as a true theory about how genetic inheritance works, rather than a purely empirical generalisation.

Thus Lewontin's conditions are not quite sufficient to guarantee evolution by natural selection. It is not enough that traits are inherited, but this inheritance must be *via* discrete units that are unaffected by the resulting phenotype: i.e. genes. Lewontin's conditions are couched in terms of patterns of inheritance observable in the traits of organisms but evolution requires that, not only does the substrate produce these patterns, but the substrate produces these patterns *in the right way*. And the 'right way' includes discrete genes that are replicated independently of phenotypic properties. Inheritance cannot be characterised purely at the level of the traits of individuals. This problem manifests itself when we try to determine whether a particular trait is *heritable*.

8.2 Nature and Nurture

Inheritance, loosely speaking, is a similarity between parents and children. However this similarity can be caused in two different ways: either because something is 'passed down' from parents to children, or because they share a common environment. Only the first form of inheritance produces natural selection, but all biological traits are the product of developmental processes that involve both. Therefore if we want to determine whether a trait will be amenable to adaptation then we must find some way of determining the contribution of genetic, rather than environmental, inheritance. This contribution is normally quantified as the *heritability* of the trait, defined as the proportion of total variance in the trait that is due to genetic, rather than environmental, variance.

There are two sets of problems in determining heritability. The first are methodological. The standard technique for measuring heritability is to raise a variety of strains in a variety of environments and then calculate their relative effects on the trait. But this technique cannot distinguish differences that are due to genetic factors from those that are due to the maternal environment provided by the uterus or the cytoplasmic environment of the zygote. For example babies born to poor, badly educated, mothers have lower birth-weight and shorter concentration spans than those born to the rich and well educated (Gonzales, 1985). Therefore, even if those children are subsequently adopted and raised in random distribution of environments — or if those children are given a common educational environment — and we subsequently find that there is a strong correlation between, say, the IQ of the birth-mother and that of the children, we still do not know

¹Thanks to John Maynard Smith, personal communication, for this point.

if this correlation is due to genetic or maternal-environmental effects (Roubertoux et al., 1990). The environment starts to affect a child well before birth, therefore environmental and genetic effects can only be experimentally separated by transplanting embryos between uteruses, or chromosomes between zygotes. Such experiments demonstrate two things. The first is that uterine and chromosomal environments do indeed have effects that would ordinarily be classed as genetic in those organisms, such as humans, where such techniques are not possible (Nosten, 1989)(Nosten & Roubertoux, 1988)(Carlier & Roubertoux, 1984) (Carlier et al., 1991). The second is that, in an operational sense, we cannot determine the heritability of a trait from the observed properties of the whole organism, but must also be able to isolate the genetic material and manipulate it directly. If there were no discrete genetic material then it would be impossible to conclusively determine the heritability of a trait using this method, and hence whether selection for it will produce adaptation.

The alternative method of determining heritability is to measure the response to selection directly: if the distribution of a trait in a population responds to selection, then that trait must be heritable². But although this is the method typically used in artificial selection experiments (Orgel, 1979), it cannot be used as a way of defining sufficient conditions for evolution since if we claim that evolution occurs when there is the differential reproduction of heritable traits, but define a trait as heritable when it evolves in response to selection, then we are left with a tautology. We need a method for determining whether a trait will respond to selection that does not reduce to observing that, in fact, it does.

The problems of separating the contributions of environmental and genetic factors are compounded when investigating human behavioural traits, since ethical considerations forbid us from manipulating environments and/or genetic material directly. For example Plomin *et al* (1977)(1985) discuss how an individual may actively or passively construct their own environment according to some genetic predisposition. Teachers, for example, may give a particularly bright child a more enriched educational environment, whilst the peers and carers of a sociable child may well reciprocate their affection. The only way to separate the effects of environmental and genetic factors in these situations would be to give a sub-population of children purposely impoverished environments, and this is not ethically possible: we would have to deliberately ignore their demands for stimulation or affection. The same confusion occurs when genetic and environmental factors are correlated for other reasons. For example in many adoption studies there is considerable selective placement, in which children are placed in homes similar to the parental environment (Loehlin, 1979). This again has the effect of confusing environmental and genetic effects.

These methodological problems in determining heritability reflect an underlying conceptual problem. The purpose of the concept of heritability is to measure the relative contributions of genetic and environmental factors to a trait. However all traits are *completely* dependent on both environment and genes. Both are 100% necessary, so does it make any sense to compare their relative contributions (Anastasi, 1958)? But although this truism applies to individuals it does not apply to differences *between* individuals in a population: if we consider a population of genetically identical individuals then the differences between them must be due to environmental factors; and if individuals are raised in the same environment then differences must be due to genetic factors. But in order to quantify heritability it is necessary to compare the contributions of each factor

²Though in section 8.4 I discuss examples of traits that *are* heritable but still do not respond to selection.

when the two effects are combined. This is done using analysis of variance (ANOVA) techniques:

Suppose we wanted to measure the heritability of the yield of wheat by raising two different strains under two fertilisation regimes³. The analysis of variance represents the causal role of these two factors by showing how the average yield of plants on a plot departs from (1) the average yield of all plants, (2) from the average yield of all plants that received as much fertiliser as those in the plot in question did, and (3) from the average yield of plants of the same strain. Suppose that the yield followed this pattern:

	Strain S1	Strain S2	Marginal Average
Fertilisation F1	1	3	2
Fertilisation F2	3	5	4
Marginal Average	2	4	3

In this case differences in strain and environment have equal effects on yield, and so its heritability will be 50%. In another case the respective yields may be

	Strain S1	Strain S2	Marginal Average
Fertilisation F1	1	5	3
Fertilisation F2	3	7	5
Marginal Average	2	6	4

in which case genetic factors have a greater effect, and so the heritability of the trait for this population in this range of environments will be greater than 50%. Moreover, in both of these cases, the combined effects of the two factors is a linear sum of the separate contributions. In the second case strain *S2* produces, on average, a yield 2 units greater than the overall mean; *F2* promotes yield by 1 unit; and the combination of *F2* and *S2* yields 3 more units than mean. However, suppose the yield is

	Strain S1	Strain S2	Marginal Average
Fertilisation F1	1	3	2
Fertilisation F2	5	9	7
Marginal Average	3	6	4.5

Here the marginal yield of *F2* is 2.5, that of *S2* is 1.5, but the combined effect of *F2* and *S2* is 4.5. The extra half a unit of yield is not due to environmental effects, nor to genetic factors, but to the interaction of them both. In other cases there may be no 'main' effects at all, with all variance due to interaction. For example, one environment may increase the yield of one strain but decrease the yield of another:

	Strain S1	Strain S2	Marginal Average
Fertilisation F1	1	5	3
Fertilisation F2	5	1	3
Marginal Average	3	3	3

In this case particular fertilisers or strains have no net effects at all. This is not a purely artificial

³This example has the same structure as human adopted twin experiments. For 'strain' read 'monozygotic twins', and for 'fertilisation regime' read 'home environment'.

example. For example Levins and Lewontin (1985) discuss two mutations in *Drosophila*, known as *Ultrabar* and *Infrabar*, which affect the number of facets in the adult's eye. *Infrabar* produces less facets than *Ultrabar* if the fly develops in an environment at 15°C, but more at 30°C. The marginal mean effects of mutations and environmental factors are very small — since they will be positive in one case and negative in another — but the combined effect is considerable. In such cases heritability gives us no useful guide as to the response of the trait to selection.

ANOVA techniques account for interaction by regarding the value of a trait as a sum of the contribution due to genetic factors, that due to environmental factors, and a third term that describes the effects of their interaction. But heritability only measures the relationship between the first two parts. This much is uncontroversial, the issue is whether it matters. Plomin argues that it does not, for two reasons.

The first reason is that Plomin claims that, as a matter of fact, the interaction effects in, for example, human behavioural traits tend to be quite small (1988, ch9). However Wahlsten (1990) points out that this is more likely to be due to the fact that ANOVA is poor at detecting interaction effects in small population sizes, rather than their absence — a problem that is more likely to occur in human adoption studies than in research on laboratory animals:

Interaction has been evaluated in studies of human IQ and usually none is seen (Plomin et al., 1977) (Plomin & DeFries, 1983) (Plomin, 1986). Generalisations have then been made that heredity and environment are truly additive, and sophisticated path models have been developed to partition variance and covariance under the assumption that interaction is negligible (Heath, 1985) (Henderson, 1982) (Phillips, Fulker, & Rose, 1987) (Plomin et al., 1985). On the other hand, an immense collection of well-controlled laboratory studies of animals has provided abundant evidence of significant and illuminating interactions between heredity and environment. At the 1987 Behavioural Genetics Association meeting in Minneapolis, the concurrent sessions on human and animal studies were like two separate worlds in terms of attitudes towards interaction. Many human geneticists dismissed interaction and cited heritability estimates with great confidence, while most of those studying mice, rats, and fruit flies documented once case of interaction after another and expressed skepticism about heritability coefficients. (Wahlsten, 1990, p112)

If the assumption of zero interaction is made and ANOVA techniques used uncritically then serious errors can occur. For example Schönemann (1997) notes that adoption experiments typically give heritabilities for human IQ of around 60–80%, which is higher than that of traits that can be more precisely measured in larger, and better controlled populations, such as body fat in pigs, or length of wool in sheep (30–50%); whilst the same measures used on such spurious human traits as wearing sunglasses after dark can give heritabilities of over 100%. Indeed, according to the same data, religious preference is more heritable than race!

However Plomin also argues that ANOVA-type heritability coefficients are meaningful even if interaction effects are present since heritability only measures the relationship between the main (additive) effects. 'Main effects and interactions are independent — mean effects of *G* and *E* are not invalidated by the presence of *G* × *E* interactions' (1990). However the existence of additive main effects is an *assumption* of the ANOVA technique, and their statistical significance in accounting for the distribution of the trait is no guarantee that they accurately reflect the underlying relationship between nature and nurture. For example, suppose that a trait were the arithmetic

product of G and E (i.e. $Y = G.E$). An ANOVA on this data may still find some additive component, but this would solely be an artefact of the method. Wahlsten (1990) (following Nabi) shows the problems that would result if the same method were used to analyse, for example, gravity. We know that the *actual* law is

$$F = \frac{Gm_1m_2}{d^2} \quad (8.1)$$

But suppose that we analysed the effects of mass (m_2) and distance (d) on the force (F) measured on an object by a fixed mass (m_1).

The experimenter's conclusions from the ANOVA would be that both mass and distance are important for the force, although the internal factor (mass) is rather more important and accounts for more variance than the external factor (distance), and mass and distance are additive because the interaction term is not even close to significance. He might even proclaim a simplified law of gravitation, $F = \mu + m + d$

It makes no sense to say that a person's weight depends more on body size than planet of residence, but this is precisely what ANOVA techniques try to measure. And there is no *a priori* reason why the interaction between genes and environment in the development of an organism will be simpler than that between mass and distance in producing gravity — indeed most of the evidence from developmental genetics points to more complexity, not less (Pritchard, 1986). Therefore it is similarly meaningless to say that a trait depends more on nature than on nurture. The analysis of variance can often succeed in hammering complex interactions into additive holes, but we should not confuse the result of this procedure with the original shape. Heritability coefficients give us a distorted picture of the relationship between genes and environment in the process of inheritance, but what is the alternative?

8.3 Inheritance and Mechanism

What do we mean by an 'inherited' trait, i.e. one that is amenable to adaptation by selection? Consider a lineage of black ravens. Their blackness may be due to genetic factors, environmental factors (such as a peculiarity in the local diet) or a combination of both. We can observe that the trait has, in fact, been inherited in the past but the question is whether it is *inheritable*, i.e. whether this trend will continue into the future. Inheritability is thus a *dispositional* property, like solubility; and just as the solubility cannot be equated with the fact of dissolving in water, so inheritability cannot be equated with the fact of having similar offspring. We need to identify some property of the offspring of the lineage that gives us some reason to believe that it will share the trait of its ancestors. We need some *justification* for an induction from past to future ravens. In section 3.5 I argued that such inductions are *not* justified simply by noting instances of black ravens, but only by explaining the observed connection between ravenhood and blackness through an understanding of the developmental processes connecting the wild-type genome of *Corvus corax* to feather pigment production. This provides good grounds for believing that all organisms that carry those genes would be black.

Inheritability is an implicitly causal property: it is not just a correlation between traits of parents and children, but rather implies that something is passed on from parents to children that

causes that trait to re-occur. If we uncover such a process then a trait will still be properly classified as inheritable even if that correlation may be disrupted by an environmental factor. Ariew (1997), for example, cites the example of opposable thumbs, which seem to be a paradigm case of an inherited adapted trait. However almost all modern humans who lack thumbs are the victims of thalidomide, and other environmental factors; therefore the variation of the trait in the population that is due to genetic effects, and hence the heritability of the trait, is very low. The reason why we are confident that thumbs are amenable to adaption is not due to heritability measures, but because we have a good idea of how they develop. In particular we know that pentadactyl limbs are the result of highly canalised processes that are (normally) very robust against environmental disruption, and so will re-appear in the lineage as soon as the environmental factor is removed. This understanding then allows us to disregard cases in which the trait fails to be passed on: thumbs are inheritable even though they have low heritability.

Of course, methodologically speaking, it is very difficult to understand the complex developmental processes that link genes to phenotypic traits; and statistical heritability experiments are a good way of identifying particular mutations worthy of further investigation. However we should not confuse the method of investigation with the object of investigation. Fisher himself — the inventor of ANOVA — warned against ‘the so-called coefficient of heritability, which I regard as one of those unfortunate short-cuts which have emerged in biometry for lack of a more thorough analysis of the data (1951, p217). And Bateson concludes that ‘analyses of statistical interaction should be the starting point of attempts to understand how developmental processes work and should not be treated as ends in themselves (1987, p2). If the environment plays no role in the development of a trait, or if its contribution is strictly separable from that of genetic factors (i.e. the variances are additive), then we can use heritability to characterise inheritance. However this is the exception rather than the rule and, in general, determining inheritability involves isolating the genetic and developmental mechanisms underlying that inheritance.

The mechanism of this genetic inheritance need not be nuclear DNA. For example Sonneborn has shown that if the patterns of cilia on the surface of *Paramecia* are surgically altered then the same pattern will be inherited by descendants (1963); moreover this transmission is independent of changes in the nucleus (1970). However the inheritance in this case is still ‘genetic’ in the sense that it is mediated by a heritable unit analysable using the same formal logic and procedures developed for DNA-based heritable systems, even though the unit in this case is composed of microtubular protein assemblies and not DNA (Frankel, 1983)(Whittle, 1983).

The same point generalises to other cases of property inheritance. For example, suppose two rich families bestow their privileges on their offspring. The first give their children a large sum of money as soon as they reach majority, but then let them fend for themselves. The second ensure that their children go to the best schools and universities, inherit titles, and are introduced to all the right social networks, but the children are never given any money directly. Both will produce long dynasties of wealth and social power. However in the first case it is the wealth that is inherited, with privilege being the likely result given a particular social environment. In the second case it is the social position that is inherited, and wealth is the result that accrues. Inherent in the concept of inheritance is the idea that something is preserved through copying, even though its effects may be dependent on the environment. The two sets of children inherited different properties from their

parents, even though the results were the same. When determining which properties are inherited it is the mechanism of the copying process that matters, not the observed result.

In the case of natural selection, correlations between parents and offspring can be achieved in one of two ways. The first is through the mechanism of a shared environment. But in order for correlation to count as inheritance there must also be some element of this causal link that is in a strong sense independent of the developmental environment, i.e. there must be an entity possessed by both parents and children as a result of which the trait may, or may not, develop, depending on the environment. Moreover, if we are to use inheritance to *explain* an observed correlation between parents and children, and hence to explain the response of the trait to selection, then that entity must have an identity independently of its role in the development of that trait.

(Of course Mendel realised all this 100 years ago when he developed the notion of the recessive gene — i.e. a gene that is possessed by an organism, and may be passed on to its offspring, but which does not have phenotypic effects. The concept of the recessive gene only makes sense if its possession conditions are independent of its expression conditions, i.e. if we can determine whether an organism possesses a gene independently of whether any observable traits are affected. However the question here is not whether natural selection is, in fact, mediated by a genetic mechanism but whether such a mechanism is a necessary condition for natural selection to proceed.)

8.4 Evolution and Development

A breed of cow that reliably gave birth to a high proportion of heifers compared to bullocks would be very profitable to farmers, but no amount of artificial selection has managed to produce one. Why not? The reason is that one more condition must be met if a trait is to respond to selection. It is not sufficient that there is a discrete genetic mechanism underlying a correlation between parents and offspring (i.e. that the trait is inheritable). It is also necessary that the *effect* of that genetic factor in the development of a trait must, to some extent, be independent of others. This is an aspect of development that Needham originally identified as ‘dissociability’:

In the development of an animal embryo, proceeding normally under optimum conditions, the fundamental processes are seen as constituting a perfectly integrated whole. They fit in with each other in such a way that the final product comes into being by means of a precise co-operation of reactions and events. But it seems to be a very important, if perhaps insufficiently appreciated, fact, that these fundamental processes are not separable only in thought; that on the contrary they can be dissociated experimentally or thrown out of gear with one another. This conception of out-of-gearishness still lacks a satisfactory name, but in the absence of better words, dissociability or disengagement will be used in what follows. It is already clear that embryonic growth can be stopped without abolishing embryonic respiration, and conversely, it is probable that growth or differentiation, under certain conditions, may proceed in the absence of the normal respiratory processes. There are many instances, again, where growth and differentiation are separable. It is as if either of these processes can be thrown out of gear at will, so that, although the mechanisms are still intact, one or other of them is acting as “layshaft” or, in engineering terms, is idling. (1933, p180-1)

If a trait is not dissociable from others then any mutation that is developmentally viable will, *ipso facto*, carry that trait. Therefore there will be no variation in that trait in the population. Even

if that trait is inheritable then it will still be impervious to adaptation by natural selection. Thus we find that no amount of selection succeeds in changing the symmetry of ocelli in *Drosophila* (Maynard Smith & Sondhi, 1960) or the sex ratios in diploid organisms (such as cattle) (Maynard Smith, 1978). Gould (1977, p234) has also used the effects of such developmental constraints to explain persistent stasis in evolution in which long-term selection fails to produce any appreciable change.

Developmental constraints can affect evolution in many ways, either by preventing any response to selection at all, or by preferring selection in one particular direction (Maynard Smith, 1985). The response of a population to selection is often likened to a ball rolling across a fitness landscape towards the valleys of optimal adaptation. However the effect of developmental mechanisms is to transform the ball into a complex polyhedron that will not necessarily roll straight down a slope but may prefer one direction to another, or even come to rest on one face.

Wagner and Altenberg (1996) discuss how such developmental constraints may be understood at the level of the genotype. They argue that a pre-condition for dissociability is the absence of ‘universal’ pleiotropy (Wright, 1968), in which every gene has an effect on every trait. If an organism is to respond to selection then its genotype must be divided into discrete ‘gene nets’ with partially independent developmental effects. If this is the case then ‘genetic change can occur in one of these gene nets without influencing the others, thereby much increasing its chance of being viable. The grouping leads to a limiting of pleiotropy and provides a way in which complex developing organisms can change in evolution’ (Bonner, 1988, p175).

If there is no genetic modularity then no amount of mutation and selection will produce reliable adaptation because selective pressures will not be able to ‘see’ — i.e. have an impact on the likelihood of replication of — individual genes. This problem can be pictured in terms of the topography of fitness landscapes. Pleiotropy increases the ruggedness of fitness landscapes since movement along every dimension in genotype space — i.e. mutation at every locus — will have an effect on the trait under selection, and hence the fitness of the phenotype. In such rugged landscapes selection will often result in populations becoming stuck on local maxima, and this is a significant practical problem in using artificial evolution as a multivariate optimisation technique (Goldberg, 1989). But if much of the genome is neutral with respect to the trait under selection then the population is less likely to be trapped and selection will be more likely to produce adaptation (Barnett, 1998) (Huynen, Stadler, & Fontana, 1996) (Kimura, 1983). Wagner and Altenberg empirically investigated the effects on evolution of genotype-phenotype maps with varying amounts on pleiotropy. They found that where development was insufficiently modular then ‘the Darwinian process of mutation, recombination and selection [was] not universally effective in improving complex systems’ (1996, p967).

Sometimes selection will not succeed in producing adaptation; and this failure can only be explained by understanding the developmental processes that produce a complete organism from an inherited genome. Dobzhansky argued that nothing in biology can be understood except in the light of evolution. But it is equally true that evolution can not be understood except in the light of developmental biology.

Genomes are not always modular, and development is not always dissociable. Pleiotropy may not be universal, but is still widespread. This fact underlies one part of Lewontin and Gould’s infamous critique of adaptationism in which they questioned the practice of

breaking an organism into unitary “traits” and proposing an adaptive story for each considered separately. Trade-offs among competing selective demands exert the only brake upon perfection; nonoptimality is thereby rendered as a result of adaptation as well. We criticise this approach and attempt to reassert a competing notion (long popular in continental Europe) that organisms must be analysed as integrated wholes, with *Bauplän* now so constrained by phyletic heritage, pathways of development, and general architecture that the constraints themselves become more interesting and more important in delimiting pathways of change than the selective force that may mediate change when it occurs. (1978, p581)

If the conditions for adaptation by natural selection are defined purely at the level of individuals, as Lewontin does, then the failure of selection to produce adaptation counts as a ‘hit’ against adaptationism. But if conditions on the mechanism of inheritance and development are included in that definition then we can account for those ‘failures’ — just as Einstein could account for the failure of Mercury to follow an elliptical orbit, and Crick and Watson could account for departures from Mendel’s second law. In this way Lewontin and Gould’s negative critique of adaptationism can be turned into a positive proposal for defining the conditions in which adaptation by natural selection occurs.

Selection for a trait will only produce adaptation to the extent that that trait is dissociable and modular; but just because a trait is dissociable does not imply that its development is independent of the rest of the organism. For example one of the most remarkable demonstrations of genetic modularity and developmental dissociability has been Halder *et al*’s (1995) success in getting complete physiological eyes to sprout on the wings, legs, and antennae of otherwise normal *Drosophila* by targeted mis-expression of the ‘eyeless’ gene. But even though the effects of ‘eyeless’ are highly modular, this does not imply that the same gene would produce a compound eye if were transcribed into, say, a mammal.

Lewontin and Gould challenged what they saw as the reductionist atomism of the adaptationist view of traits, and instead emphasised how organisms develop as whole integrated units. This is a healthy warning, but dissociability need not imply atomism. Modularity does not imply that particular genes, or gene nets, act alone, but only that changes within a gene net have limited effects on the development of other traits. Dissociability does not imply that organisms are composed of prior parts like mythological chimeras, but that those modular parts develop in, and are dependent on, the context of the whole organism. Dissociable traits are thus emergent products of whole organisms.

8.5 Conclusion

A trait will respond to natural selection only if it is inherited *via* a modular and discrete genetic mechanism, and develop dissociably. If these conditions are not met then no amount of selection will produce adaptation. Therefore a change in gene frequencies is a *necessary* condition for natural selection. But in the next chapter I will argue that a change in gene frequencies is not, in itself, a *sufficient* condition.

Chapter 9

The Role of Vehicles in Natural Selection

The organism is just the gene's way of making another gene.

— Dawkins, *The Selfish Gene*

I coined the term 'vehicle' not to praise it but to bury it. . . . The question 'what is the vehicle in this situation?' may be no more justified than 'what is the purpose of Mount Everest?'.

— Dawkins, *Burying the Vehicle*

Lewontin argues that selection can act at many levels. But selection only results in adaptation if it produces a change in gene frequencies. Gene frequencies are the bottom line of evolution. Dawkins (1976), following Williams (1966), has used this fact to argue that natural selection should be understood as a process that *only* acts on genes, rather than other levels in the natural hierarchy. In this chapter I argue that the role of organisms (and other genetic 'vehicles') cannot be so easily eliminated.

The argument for genic selectionism starts from the fact that acquired characteristics are not inherited. In its modern interpretation this has become the Central Dogma of molecular biology which states that information cannot pass from protein to DNA, but only the reverse; i.e. changes to the phenotype incurred during development will have no effect on the DNA passed on to subsequent generations through the germ-line. Therefore, as Dawkins puts it, an organism is not the object of Darwinian evolution because

to regard an organism as a replicator . . . is tantamount to a violation of the 'central dogma' of the non-inheritance of acquired characteristics. A stick insect looks like a replicator, in that we may lay out a sequence consisting of daughter, granddaughter, great-granddaughter, etc, in which each appears to be a replica of the preceding one in the series. But suppose a flaw or blemish appears somewhere in the chain, say a stick insect is unfortunate enough to lose a leg. The blemish may last for the whole of her lifetime, but it is not passed on to the next link in the chain. Errors that affect stick insects but not their genes are not perpetuated. Now lay out a parallel series consisting of a daughter's genome, granddaughter's genome, great-granddaughter's genome, etc. If a blemish appears somewhere along *this* series it will be passed on to all subsequent links in the chain. It may also be reflected in the bodies of all subsequent links in the chain, because in each generation there are causal arrows leading from genes to body.

But there is no causal arrow leading from body to genes. No part of the stick insect's phenotype is a replicator. Nor is her body as a whole. It is wrong to say that 'just as genes can pass on their structure in gene lineages, organisms can pass on their structure in organism lineages' . . .

The special status of genetic factors rather than non-genetic factors is deserved for one reason only: genetic factors replicate themselves, blemishes and all, but non-genetic factors do not. (Dawkins, 1982, p99)

This picture of evolution is summed up in Weismann's famous diagram¹ in which arrows of causal influence run solely from the genome of one generation to that of the next, with the expressed phenotype being no more than an epiphenomenal offshoot from the main branch: the organism is just the gene's way of making another gene (figure 9).

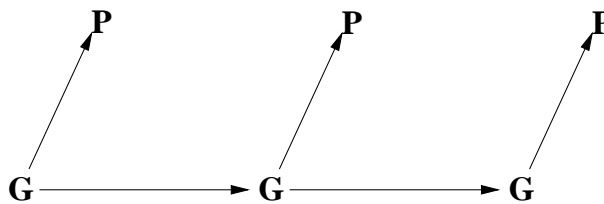


Figure 9.1:

Genes produce organisms, but genes are also *parts* of organisms. And the organism can bite back and affect the genes that they contain. This produces two kinds of problems with Weismann's diagram, depending on how those effects are understood. The first is that the replication of genes is *causally* dependent on organisms, and this dependence is the subject of the next section. The second is that the replication of genes is *conceptually* dependent on organisms, and this is the subject of sections 9.2 and 9.3. And in section 9.4 I show how we need to understand both kinds of dependency in order to understand evolution.

9.1 Burying Vehicles

Mayr (1963) describes two ways in which phenotypes causally effect the genes they contain. The first is that 'natural selection favours (or discriminates against) phenotypes, not genes or genotypes' [p184] (see also (Gould, 1978, p90) and (Brandon, 1982)). Gene frequencies change because of the effects they have on the ability of organisms to reproduce. The more successful the organism, then the more its genes will spread. The second argument is what Mayr calls 'the genetic theory of relativity', which is that 'no gene has a fixed selective value; the same gene may confer high fitness on one genetic background and be virtually lethal on another' [p296] (see also (Sober & Lewontin, 1982) and (Gould, 1978, p91)). The fitness of a gene — i.e. the rate at which its frequency changes — depends on its context (including the rest of the genome, the organism, and its environment) and is not a fixed property of the gene-in-itself.

Evolution, like all other biological processes, is a dense web of interacting causes operating at, and between, different levels of organisation. Mayr's argument is that we should pick out organisms as playing a privileged causal role in this process; but this argument stands or falls

¹This diagram has often been attributed to Weismann, though I can find no record of him actually using it.

according to how one chooses to define cause. Dawkins' defence of genic selectionism starts from the following point:

Philosophers, possibly with justification, make heavy weather of the concept of causation, but to a working biologist causation is a rather simple statistical concept. Operationally we can never demonstrate that a particular observed event *C* caused a particular result *R*, although it will often be judged highly likely. What biologists in practise usually do is to establish *statistically* that events of class *R* reliably follow events of class *C* . . . Statistical methods are designed to help us assess, to any specified level of probabilistic confidence, whether the results we obtain really indicate a causal relationship. (Dawkins, 1982, p12)

This Humean definition of cause dissolves the causal boundaries that pick out vehicles as ineliminable actors in evolution. According to this definition natural selection works *through* vehicles, rather than *on* them. To see why this is consider each of Mayr's objections in turn.

Mayr's first argument was that phenotypic properties, not genes, cause differential reproduction. Dawkins' response depends on causation being transitive. According to Dawkins' definition, if *C* causes *E1* and *E1* causes *E2* then there must be a correlation between each pair. This further implies that *E2* will be correlated with *C*, and hence that *C* causes *E2*. Now Mayr argues that phenotypic differences (*E1*) cause differential reproduction (*E2*), but those phenotypic differences (*E1*) are themselves caused by genetic differences (*C*). If causation is transitive then it is equally true that genes (*C*) cause differential reproduction (*E2*).

Mayr's second objection was that the effects of genes are context dependent. However *given* a particular context then allelic differences will be correlated with phenotypic differences (and hence with fitness). Therefore according to Dawkins' definition of cause it is still correct to say that a particular allele has a causal effect on fitness. For example, Lewontin and Sober (1982) discuss the case of heterozygote superiority in which a heterozygote (*Aa*) is fitter than either homozygote (*AA* and *aa*). In the case of human sickle-cell anaemia, for example, homozygotes for the normal allele have functional haemoglobin but are vulnerable to malaria, homozygotes for the mutant allele suffer anaemia, and heterozygotes are resistant to malaria *and* avoid anaemia. The effects of each individual allele (*A* or *a*) depends on the allele it is paired with, and so Sober and Lewontin conclude that they cannot be attributed with a unique causal role: 'if a gene raises the probability of a given phenotype in one context and lowers it another, there is no such things as the causal role that the gene has in general' (Sober, 1985, p313). However Sterelny and Kitcher (1988) argue that evolution does not require that a gene has a causal role 'in general', but only in specific contexts. A genic selectionist can perfectly well argue that an *A* allele is fitter than *a* when paired with an *a*, but less when paired with an *A*.

Both of these objections to Mayr's arguments prove only that the gene-centrist view is as valid as the vehicle-centrist and, in his first book, Dawkins compared the two points of view to the two possible views of the Necker Cube: neither is more correct than the other, they are just different views of the same process. However Dawkins later presented an additional argument that questions the validity of the vehicle-centrist view *per se*. This is the concept of the *extended phenotype* (1982, ch11–13). Consider the genes that contribute to the beaver's habit of building dams. What matters for the successful replication of these genes is not the particular actions of the beaver — the bites it takes out of trees, the driftwood it collects, and so on — but the size of the artificial

lake that it produces. The bigger the lake then the more protected the beaver will be against land-based predators and the easier it will find food. This turns the vehicle-centrist's arguments against themselves in a kind of *reductio* of Mayr's first objection: if what causally matters for evolution is the phenotypic effects of genes, not the genes themselves, then individuals do not have a privileged position in the ontology of evolution. The spread of beaver genes depends on the properties of beaver lakes, not the properties of beavers themselves. Causally speaking, individual organisms are no more, and no less, than one part of the extended phenotypic environment through which genes replicate themselves. The vehicle is buried, in Dawkins' phrase (1994).

If Dawkins (and Hume) are right about the nature of causation, then Mayr (and Lewontin) are wrong that natural selection acts on vehicles. There are two possible responses to this. The first is to argue that Dawkins is wrong about cause, as Sober does (1985). I believe that Sober's criticism is valid, but defining sufficient conditions for causation capable of supporting this claim would take us too far from the direct concerns of this thesis². Causation is a minefield that I would rather avoid if at all possible. However there is another way in which individual organisms play an ineliminable role in our understanding of natural selection. This role is *conceptual*, rather than causal, and is the subject of the rest of this chapter.

9.2 Counting Genes

What is a gene? The Mendelian answer is that a gene is a unit of heredity, but this describes what a gene *does*, not what it *is*. The biochemical answer is that a gene is a sequence of DNA, but not all possible DNA sequences are genes. So what makes a particular sequence a gene? The original answer to this question was given in the slogan 'one gene, one protein', i.e. a gene is a sequence of DNA that codes for a protein, but it is now clear that the situation is more complicated than that.

In prokaryotes, genes and the proteins they code for are co-linear; that is the sequence of amino acids in each protein is represented by a corresponding unbroken sequence of codons in the DNA. Therefore starting from the slogan 'one gene, one protein' we can clearly identify a gene as a contiguous sequence of codons on a DNA molecule. However since 1977 it has become apparent that the situation in eukaryotes is rarely that simple. Instead of forming a contiguous sequence the gene may be realised in a 'mosaic' of parts spread across the genome. The initial pre-mRNA transcript of the mosaic is then 'spliced': a process in which sections, *introns*, are edited out. The dihydrofolate reductase (DHFR) gene, for example, can vary in length from 25–31kb depending on the mammal it occurs in. Most of this variation lies in the sequence, position, and length of the introns which play no active role, however some of these introns themselves code for proteins that function independently of that coded for by the remaining exons. Moreover the exons and introns in the original DNA sequence may be spliced in many different ways depending on the actions of other regulatory genes, or on the presence of ancillary proteins. Thus two or more mosaics may intermingle across the same stretch of DNA (Breathnach & Chambon, 1981) (Wu, 1978).

Genes can also overlap on the DNA such that the same sequence of nucleotides code for more than one protein. One functional gene is often simply a truncated version of another, but in other cases the two expression processes may be read in different frames such that the divisions between

²In chapter 2 I only proposed a *necessary* condition on causal explanation (i.e. that the cause of *A* has an identity independent of *A*), and avoided the problem of sufficiency.

the codons for one gene do not coincide with those of the other. In such situations we cannot tell which amino acid — let alone which protein — a base pair codes for without observing the process of expression in action.

Genes do not even have to be tied to a particular place on a chromosome. Transposable sequences, or transposons, are capable of ‘jumping’ from one location to another either by detaching themselves from the main DNA sequence or by inducing the replication of copies that subsequently insert themselves at a new location. Most transposons are likely to be entirely ‘selfish’, with no further phenotypic effects, but it is also possible that they are a major source of effective mutations (Berg & Howe, 1989).

Genes may also be polymorphic; i.e. they may be encoded in a variety of different DNA sequences. We usually describe different sequences at a single locus as alternative alleles, but it is also possible to describe them as alternative forms of the same gene. Again the crucial factor is the effect on phenotype. There is a continuum of types of change of DNA sequences, including those that change DNA sequence but not protein sequence, those that change protein sequence without changing its secondary structure and/or function, those that create proteins with different activities, and those that create mutant proteins that are non-functional. Where on this continuum we draw the line between alternative forms of the same gene and alternative alleles depends on their role in the overall metabolism of the organism (Gusella, 1986).

In short, a gene is defined ‘semantically’ in terms of the role that it plays in the metabolism and development of an organism, not ‘syntactically’ in terms of a DNA sequence. As Lewin puts it

Genes can be isolated by working back from a protein . . . The concept of the gene itself, however, has recently evolved further. The question of what’s in a name is especially appropriate for the gene. We can no longer say that a gene is a sequence of DNA that continuously and uniquely codes for a particular protein. In situations in which a stretch of DNA is responsible for production of one particular protein, current usage regards the entire sequence of DNA, from the first point represented in the messenger RNA to the last point corresponding to its end, as comprising the “gene”, exons, introns, and all.

When the sequences representing proteins overlap or have alternative forms of expression, we may reverse the usual description of the gene. Instead of saying “one gene — one polypeptide”, we may describe the relationship as “one polypeptide — one gene”. Thus we regard the sequence actually responsible for production of the polypeptide (including introns as well as exons) as constituting the gene, while recognising that from the perspective of another protein, part of this same sequence also belongs to *its* gene. (Lewin, 1997, p146)

Suppose a copy of the DHFR gene is replicated, but in so doing mutates into one of its alternate forms. Should we describe this process as the successful replication of a single gene, or the death of an old gene and the creation of a new one? Has there been a change in gene frequencies? The answer depends on the effect of the new sequence on the organism. The concept of ‘gene frequency’, like that of ‘gene’ itself, is dependent on phenotypic properties.

Genes do not just determine the structure of proteins. Their other main function is to regulate the production of proteins, and this also raises problems in individuating genes. Sometimes a regulating gene will directly control the transcription of a neighbour. In the absence of this expressor

the regulator is functionally neutral, like a switch that is not connected to anything. In other cases the regulation is more subtle and long-range, involving the production of intermediate ancillary proteins. Gene regulation can also be affected by environmental factors that induce expression, or affect the initiation and termination of transcription (Reznikoff et al., 1985)(Platt, 1986). If the same DNA molecule is put in another organism, or *in vitro*, then different genes may be expressed.

Regulation can also occur at levels above that of a single gene. Eukaryotic genes that code for proteins whose functions are related are often organised into clusters, such as the three *lac* genes in *E. coli* which code for enzymes that decompose, transport, and aid the metabolism of, lactose. These three genes are grouped together on the genome and respond to a single regulator, forming a functionally unified *operon* that responds swiftly to the presence of lactose in the environment (Jacob & Monod, 1961). Therefore from the point of view of Mendelian analysis it is the operon, rather than its constituent genes, that is the unit of heredity.

The complexities of gene regulation mean that an apparently simple question, such as the number of genes on a genome, does not necessarily have a simple answer:

The major question about eukaryotic DNA concerns the number and types of genes in a genome. We may identify the coding potential of a genome directly, by identifying regions that have open reading frames. Large scale mapping of this nature is complicated by the fact that genes are interrupted in higher eukaryotic genomes, so that many separated open reading frames may be part of a single gene. . . . Since we do not necessarily have information about the functions of the protein products, or indeed proof that they are expressed at all, this approach is restricted to defining the *potential* of the genome. . . .

Another approach is to define the number of genes directly in terms of their expression in RNA or protein. This gives an assurance that we are dealing with *bona fide* genes that are expressed under known circumstances. It is of course the only approach that allows us to ask how many genes are expressed in a particular tissue or cell type, what variation exists in the relative levels of expression, and how many of the genes expressed in one particular cell are unique to that cell or are also expressed elsewhere. (Lewin, 1997, p645)

The first of these approaches tells us the potential capacity of the genome, but may not distinguish between functional genes, pseudogenes³, and ‘junk’ DNA. The second of these approaches can tell us more precisely how many functional genes are present, but then this number will be dependent on the cellular (and wider) environment: put the same genome in a different cell, or *in vitro*, and the number of genes changes. Gene frequencies are not determined solely by DNA sequence.

Genes, like mental states, are functional entities, defined by the relationship between the substrate in which they are realised (DNA), and its environment (the living organism). Genes are the role that DNA plays in the development of an organism, just as mental states are the role that brain states play in behaviour. If you take a DNA molecule out of its chromosomal, nuclear, cellular, and organismic environment then it does not contain any genes, just as a slice of brain tissue on a slide does not contain beliefs. Of course in order to be accorded a well-defined causal role — i.e. in order to understand how they do the job they do — it is necessary to understand how these functional genes are realised in DNA, just as we had to understand how mental representations

³Sequences of DNA that are similar to functional genes but play no active role.

were realised in brain states in order to understand how they played a causal role in our behaviour; but this should not tempt us into forgetting that it is the functional role that matters. Genes are emergent properties of the interaction between DNA and its environment. It is a truism that genes cannot replicate, or have any developmental effects, outside of the appropriate cellular and organismic environment. This is true, but is only half the story. Strictly speaking, genes do not even *exist* outside of the appropriate environment.

9.3 Counting Replicators

Natural selection necessarily involves a change in gene frequencies. But gene frequencies depend on how DNA expression is regulated by the organism. So perhaps we should do the ‘bookkeeping’⁴ for evolution in the hard currency of nucleotide sequences, rather than the fluid terms of genes? But what length of sequence? Is the unit of selection a single nucleotide, or is it the entire genome? Dawkins argues that any sequence can be a unit of selection — an ‘active replicator’ — so long as it has some effect on the phenotype, and so responds to selection (1982, p90–1).

The reason for measuring evolution in terms of arbitrary lengths of DNA rather than organism-dependent genes is that only changes in the universal currency of DNA matters for the evolutionary future. However not all DNA matters equally. Recall Dawkins’ argument why we cannot understand evolution solely in terms of the replication of organisms; namely that ‘errors that affect stick insects but not their genes are not perpetuated’. This is true, but does *not* imply that all errors that affect stick insect’s genes will be passed on either. Unless those affected genes are part of the stick insect’s germ-line cells, and unless those cells go on to form a viable zygote, then, like the stick insect’s lost leg, those genetic errors will not be passed on. If stick insects are not replicators then neither are their non-germ-line genes.

The gene-centrist view rests on a principled distinction between a replicator and the vehicle through which the replicator survives into the next generation. It is the fate of the replicator that matters, not that of the vehicle. Therefore we have to distinguish between the replication involved in vehicle-building and the replication involved in lineage-building — what Dawkins describes as germ-line, rather than dead-end replication.

The same distinction underlies the problem of ‘head-counting’ (Sober, 1985, p29). Suppose two land-dwelling organisms live on a rock surrounded by water. One has copies of gene *A* and gets fat, while the other has copies only of *B* and gets thin. The proportion of *A*-type *cells* in this two-organism population increases, but this does not constitute *adaptive* evolution: fatness does not constitute fitness. The same point also applies to the replication of sterile workers in groups of eusocial insects. The production of more workers increases the total number of the genes they carry, and may strengthen the nest, but they are as much as an evolutionary dead-end as stick-insect legs. The only germ-line replication in this case is that involved in the production of new swarming queens. In other cases there may be no prior separation between dead-end and germ-line cells. For example the apical cell of a strawberry plant runner may die after fruiting, or the runner may take root and form an offspring that lives on after the death of the original. The development of the plant thus turns a dead-end cell into a germ-line replicator.

⁴This term is due to Wimsatt (1980).

Consider another example. Many cancerous tumours are induced through the actions of proto-oncogenes that are present in the cells of the host organism and normally play a role in healthy development. The cellular proto-oncogene *c-myc*, for example, codes for a protein that helps initiate RNA transcription. These oncogenes can be implicated in the growth of cancerous tumours in two ways. The first involves the host being infected with a retrovirus, such as the mouse mammary tumour retrovirus (MMTV), which changes the way the cellular oncogene is regulated. The resulting tumour may then help the virus propagate. Or a tumour may form without viral infection due to a mutation in the regulatory genes of the mouse cell (Bishop, 1983) (Bishop, 1985) (Varmus, 1984) (Heldin & Westermark, 1984).

In both of these cases a change in the regulation of *c-myc* causes a tumour to grow. The lineage of the original cancerous cell will grow explosively, but will die with the host. But the evolutionary stories of each of these processes is very different. In the former case the replication of *myc* is a vital part of the reproductive cycle of the virus. In the latter case the cancerous replication is simply a by-product of having growth factors whose regulation is vulnerable to mutation. Thus a single act of DNA replication can be a dead-end with respect to one vehicle (the mouse) and constitute the replication of the germ-line of another (the virus).

Is cancer an adaptation and, if so, who for? In the case of mouse mammary tumours both the healthy and tumourous cells, *and* the virus, all contain the same *myc* gene. Therefore we cannot say that the cancer is an adaptation for *myc* genes *simpliciter*, but only with respect to the particular lineage of reproducing vehicles that carry it. It may be an adaptation for the *myc* of MMTV, or for the *myc* of a tumourous cell, but it is certainly not an adaptation for the *myc* of the healthy cells of the mouse. Unless we identify the vehicle through which the gene is transmitted then we have not yet answered Dawkins' question.

The point of doing our evolutionary bookkeeping in terms of DNA sequences is to determine whether evolution is occurring. But in order to do this we cannot just count the total number of copies of a sequence in the ecosystem, since it is only germ-line replication that matters for evolution. But what constitutes germ-line replication? Dawkins defines it as follows:

A germ-line replicator is a replicator that is potentially the ancestor of an indefinitely long line of descendant replicators. A gene in a gamete is a germ-line replicator. So is a gene in one of the germ-line cells of a body, a direct mitotic ancestor of a gamete. So is any gene in *Amoeba proteus*. So is an RNA molecule in one of Orgel's (1979) test-tubes. A *dead-end replicator* is a replicator which may be copied a finite number of times, giving rise to a short chain of descendents, but which is definitely not the potential ancestor of an indefinitely long line of descendents. Most of the DNA molecules in our bodies are dead-end replicators. They may be ancestors of a few dozen generators of mitotic replication, but they will definitely not be long-term ancestors. (Dawkins, 1982, p83)

The problem here lies in the words 'potential' and 'indefinite'. In one sense *all* DNA has the potential to replicate indefinitely, since we are able to extract the nucleic acid from the living organism and replicate it *in vitro*. In another sense *no* DNA can have an indefinite chain of descendents since the universe is bounded. Nonetheless the distinction between germ-line and dead-end replicators seems to have a clear biological basis; but what? Whether or not a DNA sequence

is capable of being passed on ‘indefinitely’ depends on the development and reproduction of the organism that it is part of.

Weismann (1904) believed that only germ-plasm contained the material of inheritance which was then irreversibly converted into differentiated somatic cells. According to this picture the distinction between germ-line and dead-end replication is explicit at the level of the individual cell: germ-line cells contained the material of heredity, but somatic cells did not. Weismann’s basic idea of a complete separation of germ-plasm from its phenotypic expression was absolutely correct but the mechanism he chose for making this distinction was not. We now know that *both* germ and somatic cells contain the material of heredity. Moreover some somatic cells are totipotent — i.e. they retain the capacity to produce both gametes and somatic cells. This is obvious in the case of plants which are able to grow from cuttings taken from almost anywhere on the adult. Indeed the development of cloning techniques means that, in a sense, *all* cells in *all* organisms are (potentially) totipotent. The division between germ-line and somatic cells is not as clear-cut as Weismann’s diagram implies. Whether or not a cell is destined to be a dead-end replicator, or has the potential to form part of the germ-line of a new generation, is not an intrinsic property of that cell.

Weismann pictured the relationship between development and evolution as involving a clear differentiation between somatic and germ-line cells (figure 9.2).

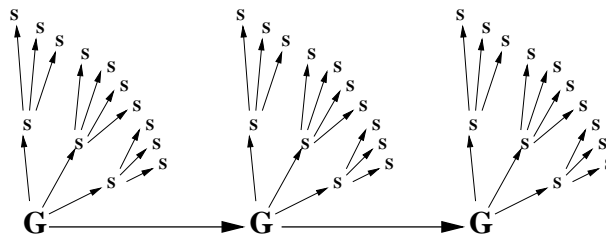


Figure 9.2:

But we now know that, in general, there is no such systematic differentiation. Whether or not a particular cell ends up as the germ-line for a new generation depends on the peculiarities of individual development (figure 9.3). Which cell ends up as the germ for a new generation is not fixed in advance⁵.

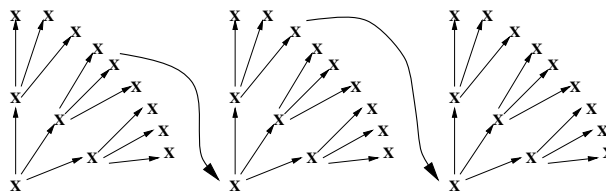


Figure 9.3:

In some situations — such as sexual reproduction — there is a more fixed differentiation between germ-line and somatic cells, and so the process of reproduction approximates more closely

⁵Dennett (1994) compares this to the ‘veil of ignorance’ in Rawls theory of justice.

to figure 9.2. But in many other cases — such as vegetative reproduction — it will not. Of course Dawkins is correct that all mutations must be instantiated in genes if they are to be inherited. But it is not determined which cells contain the genes that matter. As Buss puts it ‘knowledge of the molecular mechanics of heredity is *not* equivalent to knowledge of the units that prove heritable . . . Development controls heritability’ (1987, p14).

It is interesting to note that Weismann did not invoke his (mistaken) beliefs about germ-plasm to justify his assertion that acquired characteristics were not heritable⁶. Instead he based this argument on empirical facts about, for example, the pups of mice whose tails have been cut off, or the children of fathers who acquired duelling scars. Both of these traits were popularly believed to be inheritable — indeed I can remember being taught that Manx cats lost their tails in the same way — but Weismann demonstrated empirically that they were not. He also argued that it would be inconceivable that, for example, the well-developed fore-arms of a blacksmith could produce the appropriate changes in his sperm necessary to pass on the trait to his children. It would be akin to writing a telegram in English, sending it to China, and finding it arrived translated into Mandarin, as Weismann put it.

Weismann’s beliefs about germ-plasm stem from a consideration of development, not evolution. He argued that the development of large scale morphology in multicellular organisms could be controlled by genetic material through only two possible mechanisms. The first would be by having the entire genome copied into each new somatic cell. But then factors from outside the cell would have to ‘switch on’ the appropriate genes in the cells in all the various positions in the developing embryo: the ‘leg genes’ in the cells at the bottom, ‘liver genes’ in those in the middle, and so on. Weismann could not imagine how such a complex and delicate choreography of switches could be controlled so instead he plumped for the alternative, which was that each somatic cell would only inherit those genes necessary for their particular development. Liver cells would *only* contain liver genes, and so on. One look at a flower growing on a plant taken from a cutting should have persuaded Weismann that the former mechanism, however improbable, was responsible; but it is only relatively recently that we have started to scratch the surface of how this developmental choreography is possible.

To put it crudely, Weismann believed that given a phial of germ plasm and somatic cells then he would be able to distinguish between them. This is not true in general. Germ-line and dead-end cells contain the same DNA, and are only distinguished by the role they play in the development and reproduction of an organism. In this sense germ-line replicators are similar to straws that break camels’ backs: both are only distinguished by the role they play in the fate of a vehicle. When we look at a field of straw we cannot pick out the fatal ones. Similarly, we cannot look at a collection of cells and pick out the germ-line replicators. The object term ‘the straw that broke the camel’s back’ not only picks out a straw but also the camel for which it proved fatal; and the term ‘germ-line replicator’ not only picks out a sequence of DNA, but also the lineage of vehicles whose germ-line it is part of. In some cases, such as the *myc* gene in a tumour cell of a MMTV-infected mouse, a single DNA sequence instantiates two different germ-line replicators, each defined with respect to a different lineage of vehicles. The *myc* may be replicated either through propagation of the virus or through the spread of the tumour, and each lineage will define different adaptive

⁶Thanks to John Maynard Smith, personal communication.

pressures. In some cases the vehicle of a germ-line replicator may be what we naively think of as an individual organism. In other cases, such colonies of eusocial insects, it will be a higher group of organisms. In cases of ‘molecular drive’ (Crow, 1979) — in which genes multiply out of phase with the organisms they are in — the vehicle is lower in the hierarchy than that of the individual organism.

Dawkins *et al* argue that the bookkeeping of evolution should be done in units of germ-line replicators. But in order to pick out germ-line replicators from dead-end replicators we have to identify the vehicle that it is a germ-line replicator of⁷.

9.4 Fitness

In the last chapter I argued that the concept of ‘inheritance’, when clarified, reveals a genetic mechanism. Similarly the concept of ‘germ-line replication’, when clarified, reveals a reproducing vehicle. Therefore, although the bookkeeping for evolution may be done in the units of DNA sequences, the thing being measured is the reproduction of vehicles. Fatter is not fitter, even though it produces more copies of a gene.

This distinction between the property being measured and the units in which we measure it is explicit in the standard population geneticist’s definition of fitness:

The fitness of a particular type, *A*, is the expected number of offspring contributed by an *A* individual to the next generation. Fitness is estimated from one particular stage in the life cycle — usually the zygote — to the corresponding stage in the next. . . . Fitness is a property, not of an individual, but of a class of individuals — for example, of individuals homozygous for allele *A* at a particular locus. . . . Usually we ascribe fitness to a ‘genotype’, meaning a class of individuals with some genetic characteristic in common. . . . *Fitness is a property of a class of individuals, and not of genes.* (Maynard Smith, 1989b, p36-7, emphasis added)

Therefore when we say that a trait improves the fitness of (i.e. ‘is for the good of’) a gene, we are implicitly saying that it increases the ability of a vehicle to pass it on. Given the presence of a gene in a population the questions we must ask are not only how it aids fitness, but what vehicle it aids the fitness *of*. And the answer to this question is not always straightforward:

Genes are normally passed on by the reproduction of their vehicles, therefore fitness is usually operationally defined as the expected number of offspring of a individual member of a genotype. This definition not only covers the ‘usual’ cases of sexual or asexual reproduction of individual organisms, but also the case of transposons and other mobile DNA in which the vehicle of the gene is the DNA sequence itself: every act of replication is thus simultaneously a case of vehicle reproduction. But for the genes of infectious parasites it is not the rate of reproduction that matters *per se*, but the rate of transmission to new hosts. And parasites have evolved many ingenious mechanisms for achieving this, such as the diarrhoea induced by the cholera bacterium or the open sores produced by syphilis. Indeed in some cases too much reproduction can *reduce* the fitness of the genes of the parasite since they may kill the host before it is capable of infecting others. Therefore there is an evolutionary pressure to become *less* virulent. For example, when syphilis

⁷Indeed Hull argues that we should abandon the terms ‘organism’, ‘vehicle’, and ‘individual’ in favour of one that makes the concept of reproduction central. Thus he defines selection in terms of ‘interactors’: i.e. an entity that interacts as a cohesive whole with its environment *in such a way that reproduction is differential* (1980, p318).

first reached Europe at the end of the fifteenth century it could cover an entire body with pustules, cause flesh to fall off people's faces, and cause death within months. Now it has evolved into a much less virulent form that rarely kills the host, even if it is untreated. Therefore epidemiologists measure the fitness of such diseases in terms of the rate of transmission between host vehicles, rather than the replication of the vector or its genes (Williams & Nesse, 1991)(Ewald, 1994)⁸.

Recall that

the whole purpose of our search for a 'unit of selection' is to discover a suitable actor to play the leading role in our metaphors of purpose. We look at an adaptation and want to say, 'It is for the good of ...'. Our quest is for the right way to complete that sentence. It is widely admitted that serious error follows from the uncritical assumption that adaptations are for the good of the species. I hope to be able to show that yet other theoretical dangers, albeit lesser ones, attend the assumption that adaptations are for the good of the individual organism. I am suggesting here that, since we must speak of adaptations as being for the good of something, the correct something is the active germ-line replicator. (Dawkins, 1982, p91)

Dawkins is correct to warn against the assumption that an adaptation is good for the individual organism that carries it. However arguing instead that the adaptation is for the good of 'the germ-line replicator' does not yet answer the question, but just generalises it: germ-line replicators are always germ-line replicators *of* a lineage of vehicles, and these vehicles must be identified in order to answer the question.

In section 9.1 I argued that, although Mayr *et al* were correct to argue that individual vehicles play a causal role in the processes of selection that change the frequency of a gene, this was equally true of the rest of the extended phenotype. Therefore vehicles are not distinguished as an essential part of the characterisation of natural selection by their causal role *per se*. However in section 9.3 I argued that vehicles play an ineliminable *conceptual* role in picking out which DNA sequences count as germ-line replicators: i.e. the units in which the bookkeeping of evolution must be measured. When these two arguments — the causal and the conceptual — are put together the conclusion is that when we describe a trait as 'good for a germ-line replicator' we implicitly mean that it contributes to the ability of a *vehicle* to pass on more of its genes to the next generation. Beaver lakes can increase the frequency of beaver genes in two ways. The first is by producing more beavers. The second is by producing fatter beavers. Only the first constitutes evolution.

9.5 Conclusion

Lewontin *et al* argue that natural selection occurs whenever we have the differential reproduction of individuals on the basis of their inherited traits. This is true, but when we 'clarify' the notion of inheritance we find that correlations between parents and offspring must be mediated by a genetic mechanism. This implies that a change in gene frequencies is a necessary condition for natural selection. Dawkins argued that a change in gene frequencies is also a *sufficient* condition for natural selection. But the only changes in gene frequencies that matter are those that measure the

⁸This point will be important when we consider the analogy between infectious diseases and infectious ideas in section 10.2.

ability of a vehicle to pass on its genes to others. In the next chapter I will discuss how we can transfer this general characterisation of natural selection from natural history to social history.