Evolution of Sentience, Consciousness and Language Viewed From a Darwinian and Purposive Perspective

Chapter Seven of Nicholas Maxwell, *From The Human World in the Physical Universe: Consciousness, Free Will and Evolution,*(Rowman and Littlefield, 2001, pp. 162-201)

1 The Incomprehensibility of Double Comprehensibility

The account given in the last chapter of how free will in a physicalistic universe is *possible* also makes clear (as I have already indicated) that it is *wildly implausible*. In order for free will (or wisdom) to exist, the control structure of the brain must be such that detailed freedom-ascribing personalistic descriptions and explanations of thoughts, perceptions, desires, decisions and so on *precisely match* detailed physicalistic or causal descriptions and explanations of these same thoughts, perceptions and so on construed as brain processes, as physical processes.[1] Even though possible, it is nevertheless little short of the miraculous that such incredibly intricate matching should actually occur in the real world. It cries out for explanation and understanding.

In other words, the solution as to how it is *possible* for the human world to exist embedded in the physical universe, sketched in the last two chapters, has created a new problem (touched on towards the end of chapter six): How can the possible but wildly implausible, all but miraculous coexistence of the human and the physical be explained a nd understood?

And a little more specifically: What explanation can be given for the beautiful and intricate way in which the physical and personalistic are dovetailed together, when it comes to human beings, so as to make free will a reality in the physical universe?

In a sense, the last two chapters do not *solve* the basic problem; they transform it into the problem of explaining and understanding how the world can be such that (some) things can be explained and understood in two quite different ways simultaneously, namely: personalistically and physically. If the original mind-body problem concerns the baffling incomprehensibility of the mind, of consciousness, the new problem concerns the baffling incomprehensibility of people being simultaneously *doubly comprehensible* - *overcomprehensible* as one might say. What is incomprehensible is that there is too much comprehensibility in the world, not too little.

It is not just our brain processes that are doubly comprehensible; the whole of our world is *riddled* with double comprehensibility. This is a feature of cultural artefacts such as books, pictures, machines, musical scores, speech, drama and dance; it is a feature, in particular, of computers, robots and all control devices; it is a feature of the entire biological world; and, in a more restricted way, it is even a feature of naturally-occurring inanimate objects and phenomena such as stones, waterfalls, clouds and rainbows, in so far as these can be *described* simultaneously in terms of (a) their experiential features, and (b) their physical features.[2]

There is at once a baffling problem about the very problem itself. We have already *two* kinds of explanation and understanding available: our *new* problem is that of *explaining and understanding* the existence of that which can be explained and understood simultaneously in the *two* available ways. To what conception of "explanation and understanding" does this new problem appeal? The physical? The personalistic? Both? Or neither?[3] If a *third* kind of explanation and understanding is being appealed to, does not this just make the situation even more incomprehensible?

2 Darwin to the Rescue

In order to explain the double comprehensibility of human beings, we need to introduce two further kinds of explanation, namely purposive explanations (already briefly alluded to in the last section of chapter six), and historical explanation. The apparently *incomprehensible* double comprehensibility of people needs to be understood historically, as something that has developed gradually, as life has evolved on earth, from primitive beginnings some four billion years ago. We need to appeal to Darwin's theory of evolution, in order to render intelligible the gradual evolution of purposiveness, sentience, consciousness and free will in the physical universe.[4] As life has evolved, in accordance with the Darwinian mechanisms of reproduction, inherited variations, and natural selection, the double comprehensibility exhibited by living things has become, gradually and intelligibly, step by small step, more and more intricate, extensive and astonishing. The purposive, and ultimately the personalistic, have crept gradually into existence in the physical universe, step by understandable step.

In order to be capable of explaining the evolution of sentience, consciousness, free will, the purposive and personalistic, in this way, however, neo-Darwinism must be reformulated.

First, neo-Darwinism needs to be reinterpreted as a theory designed to explain why two kinds of description and explanation, or rather two kinds of intelligible aspects of things, the physical and the purposive, dovetail together in the intricate way in which they do, in living things. In doing this, neo-Darwinism presupposes, and exploits, both kinds of description and explanation, but in a way which does not compromise the initially *purposeless*

mechanisms of evolution (random inherited variations and natural selection) postulated by Darwinian theory. In presupposing, and being parasitic upon, other modes of explanation (the physical and purposive), neo-Darwinism is a theory very different from the theories of physics, Newtonian theory, say, or classical electrodynamics.

Second, neo-Darwinism needs to be reinterpreted so that it presupposes, and exploits, *personalistic* explanations as well as purposive and physical explanation. Once it is conceded that personalistic explanations cannot be reduced to purposive and physical explanations, it is clear that sentience, consciousness and free will cannot be accounted for by a combination of physical and purposive explanation and Darwinian theory. Darwinian theory must be reformulated, from the outset, in such a way that it presupposes and exploits personalistic explanation (where relevant) in addition to purposive and physical explanations. This does not, however, as we shall see, reduce neo-Darwinism merely to presupposing that which needs to be explained.

Third, neo-Darwinism needs to be reinterpreted in such a way that the *mechanism* of evolution itself evolves, but in a way which can be explained and understood along traditional Darwinian lines. I have already stressed, on more than one occasion, that the mechanism for evolution postulated by Darwinism (inherited variations plus natural selection) may generate purposive beings, but is itself *blind and purposeless*. This needs to be qualified. The Darwinian mechanism for evolution may *begin* as purely mechanical, blind and purposeless, but as more and more sophisticated purposeful beings evolve, capable of learning and of imitation, the very mechanism of Darwinian evolution is itself affected. Imitation, in particular, can modify the Darwinian mechanism for evolution so that it acquires a quasi-Lamarckian character, acquired characteristics (patterns of activity) being "inherited" by imitation. Gradually, step by understandable step, more and more purposiveness is incorporated into the Darwinian mechanisms responsible for evolution until, with human beings, with breeding and genetic engineering, the process can become fully purposive, or indeed conscious and intentional.

This, in outline, is what the rest of this chapter will develop in more detail.

3 Purposive and Darwinian Explanations

Purposive explanations, as I indicated in the last section of chapter six, render intelligible the actions of a goal-pursing thing, whether plant, animal, human, thermostat or robot, by explaining the actions as being designed to realize the overall goal in the given environment, but without appealing in any way to sentience or consciousness. Purposive explanations are, in a sense, degenerate personalistic explanations, devoid of the element of enabling one to know what it would be for oneself to be that robot, oak tree, or whatever. Anything that can be explained personalistically can be explained purposively, but not vice versa; and anything that can be explained purposively can, in principle (according to experiential physicalism) be explained physically, but not vice versa.[5] Personalistic explanations are compatible with, but are not reducible to, purposive explanations, which in turn are compatible with, but are not reducible to, physical explanations.

Introducing purposive explanations, in addition to personalistic explanations, might seem to make the problem of understanding the wildly improbable dovetailing together of the physical and personalistic even worse. Actually, it achieves the opposite. Purposive explanation is a denuded form of personalistic explanation, half way, as it were, between physical and personalistic explanations. In seeking to render comprehensible the coexistence of the personal and physical, it is helpful to consider the easier, but related, problem of rendering comprehensible the coexistence of the *purposive* and physical.

Although purposive explanations eschew all reference to inner sensations, experiences or states of consciousness, even when these exist, they may nevertheless refer to a sort of degenerate, functionalist version of inner experiences and mental states, such as "belief", "decision", "desire", "perception" and so on. We can say that the thermostat "believes" that the temperature of the room is 23ºC, the guided rocket "desires" or "intends" to hit its target, or "perceives" an incoming rocket and "decides" to take evasive action. We can declare that the chess-playing machine is "aware" of the threat of mate, and is "thinking" about how to meet the threat. All that is being asserted here is that the control-system of the goal-pursuing thing has states and processes associated with it that have functional roles corresponding to the functional role that real beliefs, decisions, desires etc., have in us, or in other conscious beings.

All living things are purposive. This is, indeed, *the* essential feature of living things.[6] Living things ought to be defined as naturally occurring goal-pursuing beings. *The* difference between living and non-living things is that living things are purposive in character, non-living things are not (leaving on one side non-living purposive things, such as thermostats and robots, created by us, which are, in any case, a side product of life).[7]

This, then, is the great distinction between the physical and biological sciences: biology, unlike the rest of natural science, is concerned with naturally occurring goal-pursuing things. And just as purposive explanations cannot be reduced to physical ones, so too biology cannot be reduced to chemistry and physics.

We now have three distinct kinds of explanation: physical, purposive and personalistic. Corresponding to this, we have three kinds of phenomena: the physical, the purposive and the personalistic. The physical is everything; the purposive is a (minute) subset of the physical; and the personalistic is a subset of the purposive.

We began with a problem about *possibility*. How is it *possible* for the human world, imbued with consciousness, free will, meaning and value, to exist embedded in the physical universe? This is the problem I claim to have solved (in outline) in chapters six and seven. At once, as I have already indicated, a host of further problems arise about how *in fact* diverse aspects of the human world, and the biological world, fit into the physical universe. In tackling these factual problems about how the physical, purposive and personalistic fit together, it is helpful, as I indicated in the last section of chapter six, to distinguish a number of further layers of description and explanation: (1) physical (2) molecular (3) chemical (4) neurological (5) functional, or in terms of control architecture (6) purposive (7) personalistic.

Even a complete solution to the problems of specifying precisely how the physical, the purposive and the personalistic fit together or are correlated, or how (1) to (7) are correlated, would leave unsolved the mystery alluded to above: How and why has it come to be that the physical, purposive and personalistic are correlated in the incredibly intricate, delicate way that they are, so that plants and animals can grow and act in the myriad ways in which they do in order to live and reproduce, and so that we can have consciousness and free will?

It is at this point that we need to invoke the *fourth* kind of explanation, namely historical explanation, and in particular the historical explanation of neo-Darwinism. In what follows, I indicate how neo-Darwinism needs to be reinterpreted so as to be capable, in principle at least, of providing explanations of the evolution of purposiveness, sentience, consciousness and free will in the natural world. In addition I indicate, in outline at least, how such a reinterpreted theory of evolution may actually succeed in explaining the evolution of these and other aspects of the biological and human worlds.

I shall tackle these problems of evolution in two stages. First, I consider the problem of

explaining the evolution of purposiveness in the physical universe; I then tackle the related, analogous but harder problem of explaining the evolution of the personalistic in the physical universe - the evolution of sentience, consciousness, imagination, personalistic understanding, free will, meaning, language, culture, rationality and value.

4 Neo-Darwinism and Explaining the Evolution of Purposiveness

Before I discuss the evolution of purposiveness, I want first to insert a few words about the problem of how it is *possible* for purposiveness to exist in the physical universe. So far I have just assumed that the solution to this problem is obvious: all one needs to do is to point to a simple feedback mechanism, the atom of control, of purposiveness, such as that incorporated into the most elementary thermostat
(see note 56 of the last section of chapter five). All purposive things, including those that are sentient and conscious, are able to pursue goals in their given environment in a way which is compatible with physicalism because they incorporate feedback mechanisms (in some cases hierarchical/parallel feedback mechanisms of great intricacy and sophistication, as in the conscious human brain).

This idea, despite its essential simplicity, is one of immense significance. That this is still not always appreciated, even at the dawn of the new millennium, is the outcome, I think, of a sort of historical accident. A thinker (philosopher, psychologist, engineer or biologist) might have made his or her reputation in the 19th century through expounding the idea, and spelling out its ramifications, its profound significance (for biology, for psychology, for artificial intelligence (AI), and for our understanding of ourselves). Instead the idea was discovered and developed in a somewhat scattered, piecemeal fashion, surrounded by a variety of misconceptions (from the misconception that purposiveness in biology means vitalism, or a return to Aristotelianism, to the misconception that purposiveness in psychology and AI means computing, implementing a computer programme).[8]

Once this "purposiveness via feedback and control" solution has been understood, however, it is at once clear that, in so far as it applies to life, it means that living things embody control systems of quite incredible complexity and sophistication of design. Life may be compatible with the physical: it is also wildly implausible, crying out (like consciousness and free will) for explanation and understanding.

How, then, can we explain and understand the existence of naturally-occurring purposive things, living things, in this purposeless, physicalistic universe - the coexistence of the purposive and physical? This is the problem that was solved by Charles Darwin's theory of evolution. The incredibly complex and diverse naturally occurring purposive beings that we see around us today are the outcome of a long process of evolution from a primitive beginning, evolution occurring as a result, initially at least, of the two *purposeless* mechanisms of (a) random production of inherited variation, and (b) natural selection of the variations best able to survive and reproduce.[9]

To get Darwinian evolution off the ground, we must postulate that some four billion years ago, there began to occur on earth physical processes that can be regarded as exemplifying *reproduction*, in however primitive and elementary a form. These processes might involve needle-like crystals growing in length and, at intervals, breaking into two as a result of buffeting from the environment. Postulate, next, that inherited variations can occur; in the case of the needle-like crystals, though the cross-section character of the needle is ordinarily replicated as the needle grows, nevertheless variations can occur, which are then in turn replicated along the growing needle's length. If it is in turn the case that some variations are better at growing and reproducing than others, then these are the variations that will tend to

be selected for by the environment.[10]

The essential idea is that purposiveness evolves, from meagre beginnings, as a result of the application of essentially purposeless mechanisms. The mechanisms of evolution (initially at least) are purposeless. Evolution itself is purposeless. But *what* evolves - individual living beings - are themselves purposive in character.

It is worth noting that the theory is *only* applicable to entities that can be regarded as goal-pursuing. This is because the theory requires that the entities in question can legitimately be viewed from the standpoint of pursuing the goals of survival and reproductive success. The blind evolutionary mechanisms of random inherited variations and natural selection, working together, constitute mechanisms for designing things more and more intricately and diversely adapted to pursuing reproductive success in the given environment. The first mechanism blindly produces variations, and the second tends to select those best fitted to survive and reproduce. Evolution generates diversity, including diversity of ways of living and goals; for the theory to work as an explanation, however, it is essential that throughout the diversity of goals there is a common, basic goal: reproductive success (at least in the sense of reproduction of others with genes sufficiently similar to those of the self), survival of self being a means to that end. This must be the common, basic goal of all living things (with the possible exception of human beings) if Darwinian theory is to succeed in explaining, in rendering intelligible, the evolution of purposiveness in this purposeless physicalistic universe.[11]

From the outset, in other words, the *character* of the common, basic purpose of living things plays an essential role in evolution. Darwinian evolution is only possible - or at least only makes sense - if the basic purpose of life is reproductive success. This is the only basic purpose, one might say, that has a chance of getting a grip, of coming into material existence, within a purposeless universe.

Above, I have stressed repeatedly the vital role that counterfactual truths play in purposive and personalistic explanations, and even in explanations that ascribe free will to the agent. In the case of the Darwinian goals of survival and reproductive success, the role of these counterfactual truths is strikingly apparent. Being able to act appropriately in response to a wider and wider range of possible circumstances in the given environment (thus rendering more and more counterfactual statements true) is just what a living thing needs to be able to do in order to enhance its chances of surviving and reproducing.

5 Darwin the Exemplary Philosopher

Darwin's achievement is profound. In my view it is ultimately a *philosophical* one - it being, indeed, one of the greatest contributions to philosophy. At the root of his theory is a profound, philosophical problem, generated by post-Galilean, post-Newtonian natural philosophy, (as almost all really fundamental philosophical problems are). As I have put it above: How can we explain and understand the existence of naturally-occurring purposive things in this physicalistic (and therefore ultimately purposeless) universe?[12] One is inclined to say that Darwin found the only possible solution to the problem: certainly no one has been able to think of a rival idea that is even remotely as good.[13] The basic idea is extremely simple and lucid: the ramifications of the idea are, however, immense. Darwinian evolution is the basic organizing theoretical idea behind all of biology. Furthermore, the idea has proved immensely powerful heuristically, in developing more specific theories about the features of living things, physiological and behavioral. This is what philosophy should strive to achieve: lucid solutions to conceptual problems that prove to be immensely fruitful outside philosophy itself, in science, or art, or politics, or living.

Darwin's achievement is usually, of course, understood to be scientific rather than philosophical. There are several puzzling features about Darwin's theory, however, if it is treated as a straightforward scientific theory, comparable to a theory of physics. It is quite different from any dynamic physical theory, in that it does not predict how systems evolve. Despite being profoundly explanatory, the theory does not make straightforward empirical predictions in the way in which physical theories do.[14] All physical theories have endlessly many rival (even if often less simple) theories; it is very hard even to think of a remotely plausible rival to Darwin's theory.

These puzzling features disappear the moment we regard Darwin's theory as, first and foremost, the solution to the philosophical problem of explaining the natural coexistence of the purposive and physical by explaining how and why this coexistence has evolved. In order to solve this problem, quite different from the standard problem of theoretical physics, the theory must satisfy stringent requirements very different from those required of a physical theory.

Above I pointed out how paradoxical it is to ask for an *explanation* of phenomena that can be explained simultaneously in *two* different ways; in so far as this involves employing a *third* kind of explanation, it might seem only to compound the difficulty. We are now in a position to see how Darwin's explanation cunningly avoids this problem. It does so by being a special case of the following type of *general historical explanation*, required whenever states of affairs or processes exist which are simultaneously explainable in terms of two different kinds of explanation, E_1 and E_2, compatible with one another, but neither reducible to the other. Assume that E_1 is the more inclusive kind of explanation, in that it applies to everything to which E_2 applies, and much more besides. And assume, further, that initially E_2 did not apply to anything, only E_1 being applicable to phenomena. The task of the general historical explanation is to explain how and why phenomena arose capable of being explained and understood simultaneously by means of both kinds of explanation, E_1 and E_2. (In what follows, I indicate like this, in brackets, how Darwinian theory satisfies the relevant requirement.)

The *general historical explanation* must be such that:

(a) It specifies how the phenomenon to be explained, the co-applicability of E_1 and E_2, has arisen, perhaps over a very long stretch of time. (All of life has evolved gradually from a single, primitive source during a period of some four billion years.)

(b) It may need to break the entire evolution down into many small evolutionary steps. (Big changes, such as the creation of new species, the development of multi-cell organisms, land animals or animals that can see or fly, must be shown to be the outcome of a great number of small steps, each step being selected for because it has survival-value.)

(c) These must be such that each is explainable, plausible, understandable, in both senses of explanation, E_1 and E_2, simultaneously. (Each step must be the outcome of a probable mutation (physically explicable) which must have survival-value (physically *and* purposively explicable).)

(d) A pattern of explanation must be specified for the (gradual) evolution of phenomena explicable by means of both E_1 and E_2, this pattern of explanation being such that it does not, initially at least, in any way presuppose E_2 type explanations. If this pattern of explanation does subsequently exploit a E_2 type explanation, this must be couched in terms of E_2 type phenomena whose evolution has already been explained. (The Darwinian pattern of explanation is based on replication, inherited variations, and natural selection.)

(e) This pattern of explanation must predict at least general features of the way in which E_1

and E_2 dovetail together. (Living things must be reasonably well-adapted to their environment, i.e. such as to be well-designed from the standpoint of attaining the goals of survival and reproduction. Characteristic maladaptations can also be explicable in terms of the evolutionary past.)
(f) An initial stage must be specified which is such that E_2 scarcely applies - there being hardly any trace in the events of the pattern required for E_2 to be applicable. (In the case of evolution, this has not yet been done satisfactorily.)

This special type of historical explanation for the coexistence of E_1-type and E_2-type phenomena is successful in so far as it cunningly exploits features of both E_1 and E_2 type explanations to render intelligible the gradual development of E_2-type phenomena superimposed on E_1-type phenomena. The nature of the *solution* to the problem is rigorously prescribed by the nature of the *problem*: this is why Darwin's theory of evolution seems uniquely fitted to explain evolution, and seems quite different from theories of physics.

I have assumed so far that it is proper to interpret Darwin's theory of evolution as providing an explanation for the gradual evolution of purposiveness in a purposeless universe. There is of course another view: Darwin's theory explains purposiveness away. It reveals that living things are not really purposive, and only appear to be so.[15]

One fundamental disadvantage of interpreting Darwinism in this second way is that a hiatus is artificially created when it comes to the task of explaining the evolution of genuinely purposive human life - unless one takes the heroic path of those who, like B.F. Skinner, deny that even human beings are genuinely purposive: see Skinner (1973). Adopting the view that Darwinian theory explains the gradual evolution of purposiveness in Nature (and does not merely explain apparent purposiveness away) - all living things being essentially purposive in character - creates no special, artificial hiatus when it comes to explaining and understanding the evolution of *human* purposiveness.

The basic reason, however, for rejecting the view that Darwinian theory eliminates purposiveness from Nature is that this view makes a nonsense of the theory itself. The key notion of natural selection presupposes purposiveness: natural selection involves the selection of those variations that are best able to realize the goals of survival and reproductive success in the given environment. It is only if there are entities that can be construed to be pursuing these goals, in however a primitive and elementary a way, that Darwinian theory has any application at all.

It is vital to appreciate that two notions of "purposiveness" can be distinguished, the compatibilist notion expounded above, and an incompatibilist, quasi-Aristotelian notion, which is such that, if something can be explained and understood purposively then it (or events associated with its actions) *cannot*, even in principle, be fully understood causally or physically. The incompatibilist notion, if applied to living things, implies vitalism, and a rejection of physicalism. If the *only* notion of purposiveness that we are prepared to acknowledge is the quasi-Aristotelian one, then we will be forced to interpret Darwinian theory as *eliminating* purposiveness from Nature (granted that we do not want to accept vitalism). Aristotle is a towering intellectual figure, but he should not be permitted to dominate the way we think in the 21st century. Adopting the non-Aristotelian, compatibilist notion of purposiveness makes it entirely proper, scientifically, to declare that life is essentially purposive in character, the extraordinary coexistence of the purposive and physical in living things being explained by Darwinian theory (the theory rendering the progressive evolution of purposiveness in Nature intelligible).

Those who oppose purposiveness in biology, do so under the impression that they are

opposing Aristotelianism. What the above shows is that they are themselves the *victims* of Aristotelian modes of thought. In so far as opposition to purposiveness in biology exists (due to the failure to take compatibilist purposiveness seriously), this is yet another example of the harmful influence of Aristotle on science!

Interpreting Darwinism as explaining the evolution of purposiveness in Nature has one major consequence. It means that we are justified in explaining and understanding our existence as being the (unintended) outcome of the purposive activity of animals over countless generations. There is (in principle) a purely *physical* historical explanation of our existence, it is true; but this can only explain our existence as purely *physical* beings. If we want to understand, in historical (or evolutionary) terms, our existence as *purposive* beings (necessary if we are to understand ourselves as persons), then it is both legitimate and necessary to refer to the purposive activity of our ancestors. We must refer to the skilful way that they have sought food, avoided predators, acquired mates, cared for offspring. We must refer to their "skills", their "perceptiveness", "intelligence", "perseverance", "courage", "devotion", etc., all these being interpreted in purely *purposive* ways, so that even a robot could exhibit these qualities. (Below all this will be reinterpreted *personalistically*.) Without the prior existence of all this sophisticated purposiveness, we would not exist (which is not to say, of course, that our existence was intended).

6 What Role does Purposiveness Play in Darwinian Evolution?

Darwinian evolution is not itself purposive in character; it is nevertheless the case (as I have just claimed) that the purposiveness of individual living things plays a role in evolution. Or, in other words, the purposeless mechanisms of evolution are *affected* by purposive features of the (living) things that the mechanisms operate on. The purposes of living things influence but do not determine the path of evolution.

At once the question becomes: How, and to what extent, do the purposes of living things influence evolution? Precisely what role does purpose play in evolution?

Immediately after life first began, purposiveness had, we may assume, a vanishingly small role to play in the Darwinian mechanisms of evolution. These mechanisms of random inherited variations and natural selection really were blind and purposeless. But then, as living things became, gradually, less primitive, more diversified, more diversely and richly purposeful, so that the actions of living things began to affect the survival chances of other living things in various ways, purposiveness began to play an increasingly important role in the Darwinian mechanisms of evolution, without this role being such that the mechanisms became themselves purposeful in character.

What is the minimal role for purpose in evolution? It is to contribute to the evolutionary process *only* via natural selection, the only purposefulness involved being that of the living things that are the subject of selection. What is selected for, we need to remember, is the capacity to pursue the goals of survival and reproduction in the given environment. The *goal-pursing activity* of living things (primarily *growth* in the case of plants) thus has a minimal influence on evolution from the outset, in that natural selection presupposes, and acts upon, activity (or growth) directed towards the goals of survival and reproductive success.

It is worth noting that, even when purpose plays this minimal role in evolution, it can still be a very substantial role. A change in purposive activity, brought about for whatever reason, which persists across generations, can have dramatic consequences for subsequent evolution. What has survival value (wings, feet, flippers, fur or whatever) may well depend crucially on the way of life, the kind of purposive activity that the animal in question engages in. Thus a

change in the way of life, whether brought about by a mutation, or an environmental change, or learning and imitation (to be discussed below), may change dramatically the survival value of this or that bodily feature. If a dog-like creature that runs about on land has an offspring which, as a result of a mutation, has flippers instead of legs, the outcome will be: death of the offspring. If, however, the dog-like creature, earlier, had taken to catching fish in shallow water at the edge of a river or lake, the offspring with flippers might well survive and reproduce, the outcome being the beaver.[16] Here, the flippers are caused by a mutation, but the *survival* of the creature with flippers is due to the prior change in the way of life.[17] The transition from dog to beaver is due to *two* causes; (1) a prior change in way of life, the result of individual learning and imitation, and (2) a change in genes: from the standpoint of evolution, (1) is as important as (2). In this way, changes in purposive activity may well form the leading edge of evolution. The giraffe does not acquire a long neck directly as a result of straining to reach leaves that are high up; but if the proto-giraffe was not straining to reach such leaves, the mutation which produces a longer neck would not produce an animal good at surviving and reproducing. The prior way of life is an essential part of the cause of giraffes having long necks. There is, in this wholly Darwinian and non-Lamarckian way, a partial line of causation from change in purposive activity to subsequent persistence of genetic change.[18] Changes in purposeful activity, which can be explained and understood in terms of purposive explanation, make a contribution to subsequent evolution, along with purely blind, purposeless mutations.

What is the maximal role for purpose in evolution? One extreme kind of possibility would be that the path of evolution accords with the purpose of animals, having been brought about (in part) by the active pursuit of that purpose. In other words, the end product of a stretch of evolution has come about because it was the purpose of animals that this should happen. *Our* purposes have this effect on evolution when we breed animals and plants so as to meet certain predetermined requirements, or breeding goals. The initially blind mechanisms of Darwinian evolution lend themselves to being (partially) commandeered and controlled by animal purpose in this way. This is all the more the case if new forms of life are created artificially by genetic engineering (something that has already been achieved, to a limited extent). In this case, genetic changes, instead of being blind mutations, are under purposive control, and selection, which in all other circumstances plays a vital role in evolution, can be dispensed with.

A less extreme version of breeding than the above is a common occurrence in the Natural world. Animals act as breeders, without realizing that this is what they are doing, even in a purely purposive sense of "realizing" or "not realizing" what one is doing. Sexual selection is a sort of "self" breeding. So is "offspring selection", as it ought perhaps to be called - the activity of selecting certain offspring to survive and others to die, either by killing certain offspring, or by preferential feeding and caring. "Predator" selection involves breeding living things that are increasingly difficult to eat; and "predator-avoidance" selection involves breeding predators that are ever better *as predators*. Hunting and being hunted involves mutual breeding, each breeding the other to become better and better at playing the game of hunting-and-being-hunted. Hunting leads to a kind of evolutionary arms race - a process of mutually harmful breeding of the enemy so that it becomes more and more deadly or evasive. (The breeder unknowingly breeds for the opposite of what is in his own best interests.)

The purposes of animals influence the path of evolution in other ways as well, unintended breeding being widely distributed throughout all of evolution. One might almost say: according to Darwinian theory, life breeds itself into existence without quite realizing what it

is doing. This is the *purposive core* of the theory. Indeed, we have here, perhaps, a new *formulation* of Darwinian theory, namely:
The Purposive Formulation of Darwinism: Life unknowingly breeds itself into existence.[19]

Breeding involves changing the environment of the living thing in some way, either directly by one's own actions, or indirectly. And vice versa, changing the environment of a living thing may well lead to breeding (whether intended or not). Thus changing the atmosphere, as when oxygen was formed, led to (unintended) breeding. Much more generally, whenever a living thing colonizes a new region it changes the environment of other living things in that region: this is likely to lead to breeding. The environment is significantly changed for some species if the new living things are predators, or are a source of food, or competitors for food, or the dead bodies and waste products are sources of food. Even if none of this obtains for some species in the region, nevertheless the colonizing life may still have an indirect affect, in that other living things in the region, that are predators or a source of nourishment, may be affected. Darwin himself did much to demonstrate the interconnectedness and interdependence of many different sorts living things; what this demonstrates is how widespread and important the kind of unknowing breeding that we have been considering will be in evolution. Finally, living things may influence the environment of their offspring by "choosing" (in a merely purposive sense) such and such an environment for offspring to grow up in: this too involves unknowing breeding.

Above we saw how changes in purposive action that persist across generations can have dramatic subsequent evolutionary consequences. This constitutes unknowing offspring breeding if the change in purposive activity is itself purposively explicable, as involving purposive "choice" ("choice" of new environment or food). The change in purposive activity may come about because of a mutation in genes controlling behaviour; if this leads to subsequent evolution in offspring, this only constitutes unknowing breeding of offspring in a very weak sense. (Persisting changes in behaviour do not *only* come about because of changes in genes controlling behaviour: such persisting changes may come about because of a change *of* environment, a change *in* the environment, or because of imitation.)

It is important, even at the purposive level, to distinguish:
(a) Knowing breeding: the breeder has the goal of breeding, and acts accordingly (even if not sentient or conscious).
(b) Unknowing breeding: the breeder, in pursuit of goals, acts in such a way as the actions can be construed to be those of a breeder knowingly breeding for characteristics X, even though this is not the aim of the breeder in question.

We human beings engage in knowing breeding; other creatures engage in unknowing breeding. If, however, by an "immediate-offspring" breeder we mean a parent that actively selects for such and such characteristics (related, presumably, to survival) in its offspring, then many birds and mammals that care for young may be *knowing* immediate-offspring breeders (in the special purposive sense of "knowing" being used here). Knowing immediate-offspring breeding is half way between (a) and (b).

The general picture of evolution drawn from the perspective of breeding amounts, then, to this. To begin with, breeding scarcely exists at all: purposiveness plays its minimal role in evolution. Then, as life diversifies and affects the environment, more and more *unknowing* breeding goes on. With parental care, "immediate-offspring" breeding becomes *knowing*. It is in this way that life breeds itself into existence.

So far we have discussed one general way in which purposiveness may influence evolution that is in addition to the "minimal" influence: namely via purposive activity that

can be construed to be a kind of breeding. There is another general way in which this influence may occur: namely via purposive action that directly influences the process of reproduction, and therefore what is reproduced, as we shall now see.

Interpreting Darwinism as a theory about the evolution of *purposiveness* in the world means that the theory needs to be interpreted as being primarily about evolving *ways of life, patterns of goal-pursuing activity*. In attending to evolving genes, evolving DNA, and evolving bodily structure or physiology, we are attending to that which makes possible and facilitates the way of life: but it is the way of life that is evolutionarily and biologically *fundamental*. Biology, quite properly, is primarily about *life*, about *living*, and only secondarily about *anatomy, physiology*, and biological *molecules*.[20]

In the structure of the theory, *reproduction* plays a fundamental role - in particular reproduction of *way of life*, of *pattern of purposive activity*. So far we have assumed that this is under genetic control (and influenced by the environment). But once *learning* and *imitation* come into existence, a new mechanism of reproduction comes into operation, added on to the mechanisms of genetic replication. This is *reproduction by imitation*. Thus the adult chimpanzee way of living is reproduced (1) genetically, and (2) by imitation, as a result of the young imitating actions of adults. An example of chimpanzee behaviour that is replicated by imitation rather than genetically is the action of poking a stick into termite nests to extract termites to eat; this behaviour is not, it seems, genetically controlled as some groups of chimpanzees have learned the trick, while others have not. Another well-known example is that of tits learning by imitation to peck milk bottle tops to drink the cream.

Three kinds of evolution can be distinguished, which I shall call *genetic*, *environmental* and *cultural*. An evolutionary change is (1) genetic (2) environmental or (3) cultural depending on whether it is the result of (1) genetic change (2) environmental change or (3) learning and imitation. Here, (1) and (2) are involved from the outset of evolution, but (3) is a relatively recent form of evolution. In fact (3) is a special form of (2).

There is a crucial terminological point that I must now stress. The term "cultural evolution" is usually taken to mean (a) "the evolution of culture": see, for example Dawkins' discussion of "memes" (Dawkins, 1978, ch. 11); or Boyd and Richerson's book *Culture and the Evolutionary Process* (1985). Here, however, by "cultural evolution" I mean (b) evolution of living things when the *method of reproduction* involves not just genes, but learning and imitation. "Cultural evolution", here, means "evolution (in part) by cultural means". According to (b), then, *what* evolves is exactly the same as before, individual living things pursuing characteristic ways of life; it is the *method* of reproduction that is new, in that it includes a cultural component, a component of learning and imitation. (In what follows, risking tedious repetitiveness, I shall persistently emphasize that "cultural evolution" here means "evolution by cultural means".)

All that I require for what follows is that (b) is an entirely legitimate way of construing "cultural evolution" within Darwinian theory: I do not need to establish that it is the *only* legitimate way, and that all talk of "evolving memes" is illegitimate.

Nevertheless, it does seem to me that (b) - evolution by cultural means - should be taken as the primary, proper way to interpret and apply Darwinian theory in connection with purposive life that begins to evolve by means of cultural development. This accords with the overall view, argued for above, that Darwinism should be interpreted as being about the evolution of purposiveness in nature. It enables us to see what is special about human evolution - the dramatic flourishing of "cultural evolution" - as something that has deep roots in conventional Darwinian evolutionary processes, as we shall see below.

There are, furthermore, serious difficulties connected with construing "cultural evolution" in the way proposed by Dawkins (1978, ch. 11), and subsequently defended by Dennett (1996, ch. 12), as the evolution of a new kind of entity, the meme, by means of Darwinian mechanisms. As David Holdcroft and Harry Lewis (2000) have pointed out, there are too many disanalogies between evolving life and evolving memes. It is not clear how the individual meme is to be identified; nothing seems to correspond to the distinction between gene and phenotype (roughly, the body of a living thing); meme evolution continually involves lines of descent separating and subsequently merging, in a way that does not happen in ordinary biological evolution; it is not even clear what a meme *is* (i.e. whether it is an artefact, a pattern of activity, an idea, or a pattern of neurological processes in the brain[21]). As it happens, both Dawkins and Dennett acknowledge such disanalogies between the evolution of memes and the evolution of life. At most, Holdcroft and Lewis conclude, the whole idea of memes evolving in a Darwinian way amounts to no more than a metaphor: taken literally, as a theory of the Darwinian evolution of memes, it is false.[22]

There is an even more important point. Tackling "cultural evolution" in the way I advocate, as "evolution by cultural means", provides a framework (the standard Darwinian framework) within which the gradual evolution of aspects of culture, such as meaning and knowledge, can be understood: see, for example, the discussion of the gradual evolution of meaning below. This is because this approach automatically relates cultural artefacts to purposeful (and even consciously intentional) *action*.[23] The meme perspective renders this impossible, and thus obscures just that which needs clarification. Meaning, knowledge and communication begin before cultural evolution gets under way, and thus before memes exist (see section 8, vii, below): the meme viewpoint thus cannot account satisfactorily for the evolution of meaning and knowledge.

One can understand why Dawkins should have opted for the meme perspective. Having so brilliantly argued for a gene-centred perspective, genes selfishly manipulating phenotypes they inhabit to produce more copies of themselves, one can see how and why Dawkins should come to see cultural entities as acting in the same way, manipulating the brains they inhabit to produce more copies of themselves. But better, by far, to interpret Darwinism as being about evolving goal-pursuing living things, mechanisms of reproduction evolving as evolution proceeds.

It is tempting to think that memes might be the genes of cultural evolution: what genes are to genetic evolution (a part of the mechanism of reproduction of purposive living things), so memes are to cultural evolution (a part of the mechanism of reproduction of purposive living things, evolving by cultural means). But this is not quite right either. It would be more correct (but still not quite right) to say that it is *the instinct and capacity to imitate plus the meme* that is the gene of cultural evolution. In the end the most that can be said is that, with the emergence of culture, of something like memes in evolution, it is the method of reproduction that is new, not *what evolves* (as Dawkins and Dennett hold).

As I have remarked above, cultural evolution (evolution by cultural means) is quasi-Lamarckian in character, in that it involves something like the inheritance of acquired characteristics, namely: "inheritance by imitation" of purposive activity learned by parents (or ancestors).[24] Purposive activity is handed on across generations *directly*, only the *capacity to learn and imitate* being reproduced genetically: it is this which makes it possible for cultural evolution to be cumulative, rapid, and even under purposive control. Evolution by cultural means is profoundly important for our understanding of our own existence, for we are, to a unique extent, the product of a long, massive process of cultural evolution (interacting with

genetic and environmental evolution). Thus, in order to understand the evolution of human consciousness and free will, it is important to understand the Darwinian account of the evolution of cultural evolution. What are the Darwinian reasons for cultural evolution (evolution by cultural means) to come into existence?

There are two components to cultural evolution: the capacity of individuals to learn to do new things; and the capacity to imitate (that is, to learn from the learning of others).

The first capacity has obvious survival value; its widespread existence in the living world can thus be understood along Darwinian lines. (It deserves to be noted in passing that the coming into existence of the capacity to learn in itself has consequences for subsequent evolution: being able to learn means being able to live in more diverse ways, in more diverse environmental niches, which in turn creates more diverse kinds of survival pressures, this in turn making subsequent evolution into diversity more likely.)

The second capacity (being able to imitate) is of survival value especially when there is parental care, since imitating parents probably has survival value (since if one's parents were not good at surviving and reproducing one would not exist oneself). Parental care, in turn, is one of two basic strategies for reproducing: production of vast quantities of potential offspring, or parental care of a few potential offspring. The second has a long evolutionary history: crocodiles and birds exercise parental care, and it is reasonable to suppose that dinosaurs exercised parental care. (It would be interesting to know to what extent parental care is itself, in birds, and in chimpanzees, the outcome of imitation.)

The special feature of cultural evolution (evolution by cultural means) is that it can be cumulative across generations (each generation building or elaborating on what has gone before) without it being necessary to wait for the blind vagaries of mutations to produce what is required (if they ever do). Above all, cultural evolution (evolution by cultural means) makes possible the progressive creation of language (without the absurdity of each linguistic development only being made possible by the appropriate, prior genetic change).

Once cultural evolution has come into existence it becomes possible for aspects of evolution to come under *purposive* control. A project can be adopted, with a long-term aim, that is handed on from generation to generation by imitation. The development of speech and writing enormously enhances this process. The history of endeavours that persist across generations, whether successful like science, or somewhat less successful like the search for world civilization, becomes possible. Genetic plus environmental plus cultural evolution eventually becomes history (so far, of course, only in a purposive sense).

Cultural evolution can have dramatic consequences for genetic and environmental evolution. The beginnings of a language arising as a result of cultural evolution may well lead to a situation which is such that being good at using the language is essential for reproductive success: if you cannot speak well, you cannot get a mate. A new environment is created such that, being-good-at-speaking is selected for. Our capacity for language acquisition and use, genetically programmed, has almost certainly arisen in this sort of way, as a result of the prior existence of language created by cultural evolution.[25] Cultural evolution leads to genetic evolution, to changes in physiology.[26]

This concludes my account of the way in which the Darwinian explanation of the evolution of purposiveness itself subtly, and entirely legitimately, uses the concept of purpose - evolution itself being under changing and increasing purposive influence, to the extent that the whole character of evolution is gradually changed. It is vital to appreciate, however, that the concept of purpose that is being appealed to here is the compatibilist, anti-Aristotelian, non-reductionist concept, indicated above. As life evolves, the role for purposive explanation

becomes greater and richer; being compatibilist, however, it does not negate the possibility of everything being explained physically; but equally, being non-reductionist, purposive explanations are an essential ingredient of an historical, evolutionary understanding of life today, *necessary* if we are to understand ourselves, and other living things, as purposive beings. A purely physical account of evolution could only enable us to understand ourselves as *physical systems*, not as *purposive beings*.

7 Is an Evolutionary Explanation of Sentience, Consciousness and Free Will Possible?

So far we have discussed how Darwinian theory is able to give an evolutionary explanation for the coexistence of the purposive and physical in the world. Can we, analogously, show how Darwinian theory is able to give an evolutionary explanation for the coexistence of the *personalistic* and the physical in the world? Can neo-Darwinism explain the evolution of sentience, consciousness and free will?

At once a difficulty arises. Darwinian theory is concerned with reproductive success; but reproductive success depends on what you *do*, and not on what you *think or feel*. It may be that sentience, consciousness, free will, play a vital role in producing action required for survival and reproductive success: but if so, all that is required is "sentience", "consciousness" and "free will" in a thoroughly *purposive* (or functional) sense of these terms. As long as the relevant people or animals are behaving *as if* they are sentient, conscious or free, it cannot matter at all that they are not *really* sentient, conscious or free. A Darwinian explanation of the *personal* (as opposed to the purposive) seems, in other words, impossible.

This may well appear to be *the* problem confronting the development of a Darwinian explanation of consciousness. Given the viewpoint developed in this and the last two chapters, however, this problem disappears. Darwinian explanation, like historical explanation generally (of the type specified above) has the task of explaining the coexistence of two kinds of comprehensibility, E_1 and E_2. It uses, subtly, both kinds of explanation. So far, E_1 and E_2 have been physical and purposive explanations respectively. We have already agreed, however, that *personalistic* explanation cannot be reduced to purposive or physical explanation, or any combination of the two. Therefore, as long as Darwinian explanation is concocted in terms of physical and purposive modes of explanation, this can never, of itself, give rise to *personalistic* explanation, or explain how and why the *personalistic* (sentience, consciousness, free will) comes into existence. If our Darwinian explanation is to succeed in rendering intelligible the coexistence of the physical and the personalistic, then we must feed in the personalistic mode of explanation from the outset, as it were; we cannot extract it from a purely physical and purposive starting point.

Granted that we have sentient, conscious, free, or even wise life before us, and granted that this is adequately described, explained and understood *purposively*, what does the corresponding *personalistic* explanation add to this? It tells us what the creatures or persons experience, feel, see, think, desire, fear, decide, imagine *as we ourselves might experience, feel, etc., these things*. The purposive account is severely impersonal in character; it treats sentient, conscious beings as if they were robots, merely simulating the having of experiences, feelings, perceptions and so on. Personalistic understanding of the mouse, the bat, the owl, the fox or chimpanzee enables us to understand what it feels like to be the mouse, the bat, the owl, etc. The myriad smells of the night air waft before us; the rustle of leaves, the flickering of a shadow, the echo of our squeaking, the sharp pang of hunger, the confused pleasures and pains of animal life, come vividly into existence in our imagination.

Personalistic explanations will be functionally similar to corresponding purposive explanations, but will be richer in content in that the experiential dimension will be added on. There will be meaning and value, contentment and tragedy, pleasure and anguish, and reasons for action that only those who have had the appropriate experiences, desires and fears can understand.

Adding on the experiential dimension might contribute to understanding why certain sorts of sensations or perceptions have more survival value than others. Consider, for example, the evolution of the perception of colour. At an early stage, we might imagine, perceived colours are such that they do not sharply differentiate differently coloured objects, making ripe fruit or predators that much more difficult to discern.

Gradually, as the brain evolves, perceived colours become more and more vividly distinct. An experiential understanding of why one set of colour-experiences has greater survival value than another set becomes (in principle) possible, even though this will be mirrored by a purposive (or functional) understanding, which will not require that one has the relevant colour experiences oneself.

To sum up, personalistic/physical evolutionary explanations will match, but will not be reducible to, corresponding purposive/physical evolutionary explanations.

8 The Evolution of Sentience, Consciousness and Free Will

We come, at last, to the task seeing how it might be possible for neo-Darwinism to explain the evolution of sentience, consciousness and free will. Of these, the evolution of sentience presents, in my view, the greatest mystery. Once inner experiences of some kind exist - sensations of pain and pleasure, tactile, visual, olfactory or auditory sensations - it is just about conceivable that further elaboration of such inner experience becomes consciousness, self-consciousness, the full range of human awareness. But the first step, the beginnings of an inner world known to itself, seems inherently inexplicable.

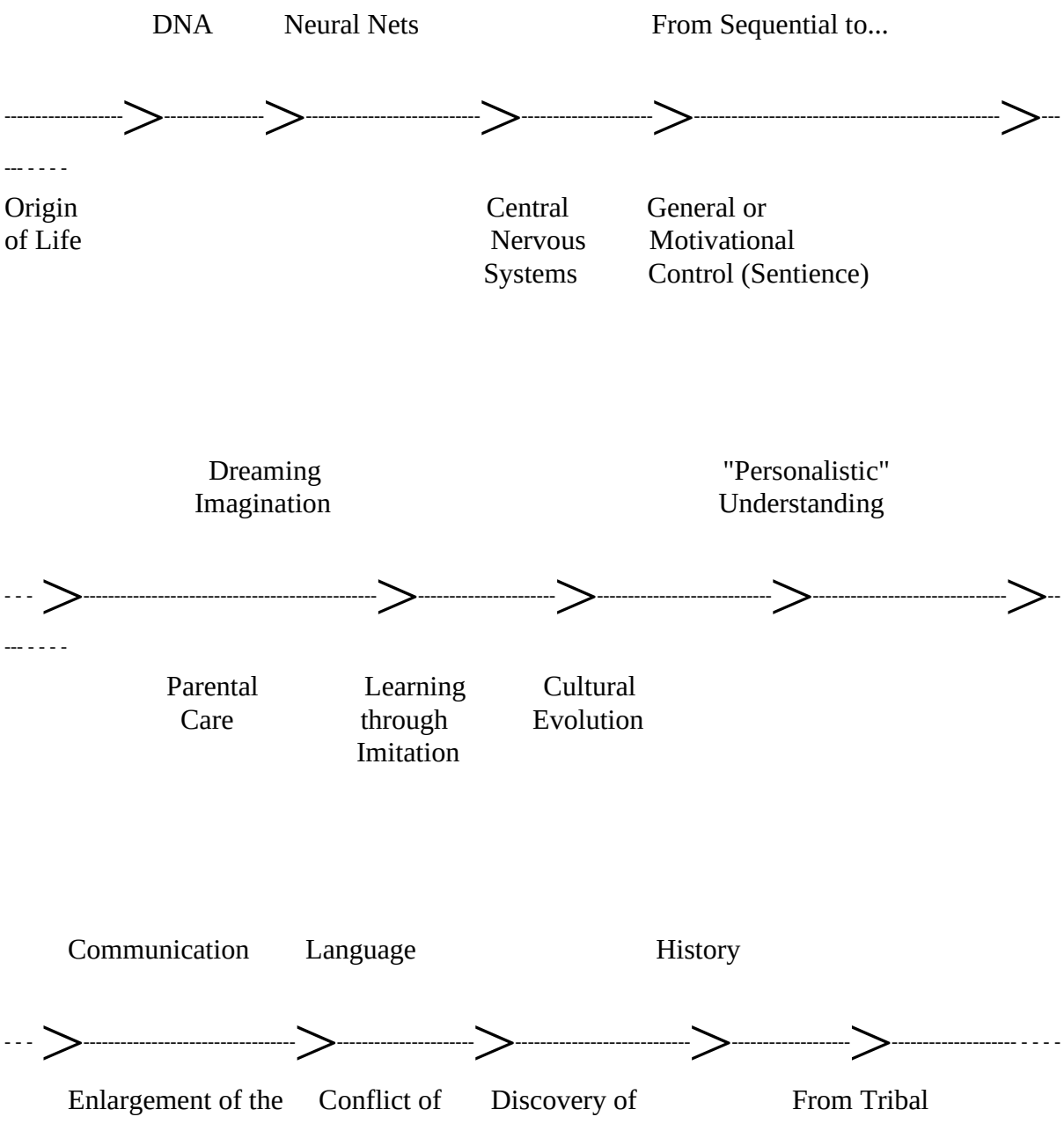
In part, of course, as we have seen above, the sense that there is a profound, unsolvable mystery here is based on a misunderstanding. It is based on the idea that a physical-plus-purposive evolutionary explanation ought to suffice to explain the emergence of sentience, and yet seems hopelessly ill-equipped to do so. Once it is appreciated that physical-plus-purposive descriptions and explanations cannot in principle capture or encompass the experiential, the sentient, for reasons given above, it becomes clear that in demanding this we are demanding the inherently impossible. What we can legitimately set out to discover is how sentience, at its most primitive, *correlates* with purposiveness, with control or functional aspects of the brain. What developments in brain structure and function are in fact associated with the first primitive beginnings of sentience? (This question, though legitimate to ask, is nevertheless fiendishly difficult to answer, due to the inherent difficulty of determining whether such and such a living brain, of a sparrow, let us say, or a mouse, really does have sentience associated with it. In order to compare and contrast mouse and human sentience we require a composite mouse-human brain, able to generate neurological processes associated with both mouse and human sensations. But would even this bizarre possibility tell us what we want to know? Mouse-type neurological processes might give rise to sentience in the mouse-human brain, and yet give rise to no sentience at all in the exclusively mouse brain.)

In what follows, I argue that human consciousness and free will have arisen as a consequence of a number of prior, cumulative evolutionary steps, involving the emergence of such things as motivational control, learning, imagination, imitation, personalistic understanding, intentional communication. Each step *may* be interpreted in a purely purposive way; indeed, even consciousness and free will may be so interpreted: my argument

is, however, that as these control, purposive steps are reached so genuine personalistic sentience, consciousness and free will come gradually into existence over evolutionary time. In what follows I do not simply postulate what I seek to explain; rather, I postulate the emergence of a new control capacity which is, I argue, correlated with the emergence of the experiential and personalistic. The succession of evolutionary steps[27] towards human consciousness that I wish to stress is summed up in the diagram.

A precondition for the emergence of sentience is, I suggest, the evolution of brains that (a) control an increasing variety of actions and (b) do so increasingly by means of what may be called *active* control as opposed to *sequential* control. Active control has a kind of flexibility,

Diagram Steps in the Evolution of Consciousness and Free Will



Imagining = Making neurological processes occur in your brain that are analogous to the processes that would occur were you actually doing what you imagine you are doing.

Personalistic Understanding = Imitating in imagination what the other person is doing and imagining.

Imitation + Imagination = Personalistic Understanding

a capacity for learning, which sequential control does not have; thus in linking the emergence of sentience to the emergence of more and more active control over a greater and greater variety of actions, I am linking it to the capacity to *learn*.

The functioning brain (whether of insect, mammal or robot) can be thought of as consisting of three interacting parts: a *goal-representing* part, G, which represents the goal that is actively being sought; a *perceptual* part, P, which represents the current state with respect to the goal, and a *motor* part, M, which controls movement so that the current state moves towards attainment of the goal. As time passes, the animal or robot will pursue a succession of goals, g_1 ... g_n. It may be that a properly completed succession of such goals adds up to the attainment of some overall goal, g(1...n). Granted that the animal or robot pursues g(1...n) by pursuing g_1 ... g_n in proper sequence, there are two quite different kinds of control-structure which make this possible.

i Sequential Control. At any given instant, the actively controlling part of the brain, G+P+M, is pursuing one or other of g_1, g_2 ... g_n. Furthermore, the brain is so structured that completion of g_1 triggers G to represent g_2, completion of g_2 triggers G to represent g_3, and so on until g_n (and hence g(1...n)) is attained.

ii Active Control. At any given instant, the actively controlling part of the brain, G+P+M is such that G represents the overall goal to be attained, g(1...n), G+P+M operating so that g_1, g_2 ... g_n are in turn pursued and attained.

The actively controlling brain represents, in G, the basic goal g(1 ... n) being pursued, and adjusts M in the light of P so that g(1 ... n) may be attained. The sequentially controlling brain represents, in G, in sequence (if all goes well) g_1, g_2 ... g_n, at any stage P and M operating so that the goal of that stage may be attained.

Sentience arises, I suggest, as the variety of goals sought increases, and as sequential control becomes, more and more, active control.

Consider the following example of sequential control. There is a kind of wasp that lays its eggs in a burrow. The wasp first builds the burrow, then flies around looking for a cricket which is then stung, paralysed, and taken to the burrow. The wasp deposits the cricket outside the burrow, inspects the burrow, emerges to drag the cricket into the burrow, lays the eggs, and closes the burrow. When the eggs hatch, the paralysed but still living cricket is food for hatched grubs.

On the face of it, this looks as if the basic goal was under "active control" and, to that extent, the wasp "knew what it was doing". This is, however, refuted by the following experimental result. If, when the wasp is inspecting the burrow, the cricket is removed a little from the burrow, the wasp will emerge, move the cricket to the predetermined distance from the burrow, and reinspect the burrow; and this can be repeated up to forty times.[28]

The conclusion is clear. The wasp is operating under sequential control. Its brain actively pursues in turn something like the following goals: g_1: build burrow; g_2: find, sting and paralyse cricket; g_3 return to burrow and place cricket 5 cm. from the burrow; g_4: inspect burrow; g_5: drag cricket into burrow; g_7 close burrow. The completion of one goal triggers the representation and pursuit of the next. The overall way the subordinate goals are pursued and put together cannot be adjusted by what is learned during the pursuit of any subordinate goal.

It is just this latter that is possible when the brain actively controls the pursuit of the basic goal. In this case, something that is learned during the pursuit of a subordinate goal, g_r, *can* affect the way all the other subordinate goals are pursued and put together in time. Active, functioning feedback mechanisms relate to the overall goal and not just to the subordinate goal being pursued.

There is, I suggest, a further step involved in the emergence of sentience, namely, the emergence of:

iii Motivational Control. The more it is the case that quite complex sequences of actions leading up to the attainment of a goal - such as hunting - are under active control, the more problematic becomes the means whereby such control is to be achieved (as far as brain-design is concerned). At a certain stage this problem is solved by active control becoming what may be called *motivational* control: the goal is represented in the goal-determining part of the brain, G, in such a way that the animal is prompted to try to attain the goal *without any definite strategies for obtaining the goal being preprogrammed in the brain*. The animal has to learn, to discover for itself, as it were, how the goal is to be achieved. The animal is prompted to pursue a goal specified in only rather general terms by neurological processes going on in the animal's brain; precisely what goal is to be pursued, and what methods are to be adopted, has to be "worked out" by the animal, and will depend on past experience, on what has been learned. The goal, here, might be food; in this case, neurological processes go on in the brain which prompt the search for food without any specific sequence of actions being predetermined. This, I suggest, is the basic requirement for the emergence of sentience; in the particular case considered, the neurological processes that prompt the search for food would be experienced by the animal as *hunger*.

On this view, then, neurological processes begin to be experienced as *feelings* and *desires* when they are such that they prompt the animal to act in ways that are determined in only rather open-ended, non-specific ways, this making flexibility and learning possible. The animal feels hunger, thirst, fear, anger, exhaustion, curiosity, sexual attraction, protectiveness towards young, and acts accordingly. Animals do what they feel like doing, what their desires and fears prompt them to do, and natural selection ensures that their brains are so designed that, in the given environment, feelings and desires, or, in other words, the corresponding control processes in the brain, prompt animals to do what they need to do to survive and reproduce. Furthermore, *perception* becomes sentient as a result of being linked to motivational control, to feeling and desire.[29]

Spiders do not build webs as a result of a generalized, insistent impulse to build a web; they are under precise, sequential control to execute a series of quite specific actions, which leads to the construction of the web. There is no sentience, no feeling or desire, no sensation, and no (or little) learning. Lions, on the other hand, I surmise, do feel hunger; they do experience a generalized, insistent impulse to find food. What needs to be done to obtain food, namely hunt, has to be learned. Nature even arranges for hunting skills to be acquired during youth: cubs experience the irresistible impulse to engage in play, in mock combat.

Unlike spiders, lions do have feelings and desires, and do experience sensations.

In short, as brains, during the course of evolution, become sufficiently sophisticated to guide the animal to pursue diverse goals of life necessary for survival and reproduction by means of motivational control (construed in purely purposive terms), so authentic *sentience* comes into existence, the genuine experiencing of sensations, feelings and desires. Motivational control, and hence sentience, has survival value, and is thus selected for, because it facilitates learning, and learning has survival value.

We can see, further, how motivational control can lead to the beginnings of:

iv Choice and Free Will. Once a brain is sufficiently sophisticated to work by means of motivational control, it does not need to be very much more sophisticated to be able to represent more than one desire or feeling simultaneously, the stronger (or strongest) being the one that is acted upon. Simultaneous experiencing of conflicting feelings or desires, such as hunger and fear, may however enable animals to learn, to improve their ability to act, in circumstances of conflict, as when there is both hunger and danger. Animals constantly find themselves in such situations of conflict, in situations which require conflicting modes of behaviour, such as exploration for food, and flight from a predator. It is conceivable that simultaneous representation of the corresponding conflicting feelings and desires, as a part of the operations of motivational control, facilitates *learning* how to act in such a way that the best justice is done to the two conflicting desiderata, in particular *learning* how to switch rapidly from one mode of action to another, when such a switch is required. Thus the mouse learns to venture forth cautiously from its hole when there are indications of the presence of cat, and learns when, with too great proximity of cat, to flee back into the safety of its hole. From *find food* and *avoid cat* the mouse learns to pursue a new goal: *find food in such a way as to avoid cat*.

This learned weighing of contradictory desiderata, this creation of new goals that best reconcile given contradictory goals, vital for survival, made possible by motivational control, may be thought of as the evolutionary origins of human choice and free will.

The fundamental goal of the mouse (survival and reproductive success) controls the learning of how to reconcile in practice the contradictory desiderata of food and safety. Human free will differs from this in part because humans have incredibly diverse, complex and variable aims and circumstances in comparison to those of the mouse (although the mouse might view the matter differently!). There is, furthermore, the difference that the fundamental life-goal (or life-value) of the person is open to a much wider and richer range of interpretations than the fundamental life-goal of the mouse, as specified by Darwinism. And there is the difference that the fundamental life-goal of the person may evolve with time in a way which does not happen with the fundamental Darwinian life-goal of the mouse. (And there are other differences, such as differences in consciousness.) What mouse-choice and human-free-choice (or even human-wise-choice) have in common, however, is the presence of *learning* how to choose between conflicting desiderata, this *learning* going on in the interests of attaining the fundamental goal (or value) in life.

In section 12 of the last chapter, we saw that learning to acquire greater free will, learning to become wiser, ought to form an integral part of the concept of free will or wisdom, and helps solve the problem of the flexibility, the lack of rigidity, intuitively felt to be inherent in the notions of free will and wisdom. The considerations just developed about mice and men suggest that "being able to learn how to make better choices between goals" ought perhaps to be regarded as basic to the difference between "free choosing" and mere "switching".

The next evolutionary step in the creation of consciousness is the development of:

v Imagination. Imagination arises whenever a control system is sufficiently sophisticated to produce control-processes analogous to control-processes (perceptual and motor) that would be going on were the entity (animal, person or robot) actually to be performing some action in some environment, such as, for example, climbing a tree. In this particular case, the entity "imagines" that it is climbing a tree.

This is a purely purposive, functional characterization of imagination. If, however, the control system, the brain, operates sufficiently by means of motivational control for the purposive entity to be sentient, then "imagination" as just characterized becomes authentic imagination, the genuine creation of inner experiences of possible states of affairs, more or less as we know it.

It is conceivable that, from an evolutionary and neurological standpoint, imagination grows out of motivational control. The lion's *feeling* of hunger (persisting neurological processes) engenders the occurrence of neurological processes analogous (in control terms) to those that would occur were the animal to be hungrily devouring a freshly killed deer, and this in turn suggests and prompts the initiation of appropriate action, in this case hunting. Motivational control may require something like imagination if it is to become operable.

If this were the case, imagination would have obvious survival value. But, quite apart from this possibility, being able to imagine has obvious survival value - for example, for animals that hunt. Suppose that lions are hunting deer. Three or four possible strategies are available: charge straight down the hill at the deer; crawl slowly towards the deer through the grass; send one or two lions to the other side of the deer, and chase the deer towards the ambush. If lions are unable to imagine, and are obliged to try out these options in reality, attempting to put the first option into practice may scatter the deer and destroy the possibility of a successful hunt. But if the lions are able to imagine, they can try out options *in the imagination*, failure in the imagination leaving the situation undisturbed. Trying out possible solutions to problems in the imagination is always going to be less costly than trying out possible solutions in reality: in the extreme case, it spells the difference between life and death - dying in imagination being very much less lethal than dying in reality.

The achievements of imagination can be so impressive that some psychologists have been misled into thinking that an entirely new faculty of "insight" is required to explain them. Take the case of Kohler's experiments with apes.[30] In one such experiment, the ape is presented with bananas hanging in the cage out of reach, and two boxes which, if piled one on top of the other, will enable the ape to reach the bananas. An ape without "insight" might blunder around, stupidly trying out all sorts of solutions until the correct solution is hit upon almost by accident. An ape ostensibly with "insight", sits and ponders until, suddenly, he gets up, piles up the boxes and gets the bananas. The second ape appears to be vastly more intelligent than the first. Actually, the *only* difference may be that the second ape blunders around, stupidly trying out all sorts of possibilities *in the imagination*. The second ape is just as stupid as the first, but keeps its stupid efforts at solving the problem hidden from view by trying them out in the imagination only. There is no need, as Kohler thought there was, to postulate some additional, mysterious faculty of "insight". (Or, alternatively, insight amounts to nothing more than being able to solve problems, perhaps very stupidly, in the imagination.)

Dreaming is a kind of imagining. Indeed, one might almost characterize imagining as dreaming when you are awake. The fact that mammals dream is a strong indication that they can imagine.

There is here a possible Darwinian explanation for dreaming. The survival value of dreaming is rather generally regarded as something of a mystery. It may be, however, that

we dream in order to develop the capacity to imagine. Dreaming has survival value because imagining has survival value.

The consequences of being able to imagine are immense. It vastly increases the scope of the arena within which one acts, both in space and time. If you cannot imagine, you cannot imagine that you are at other places and times, you cannot conceive of other places and times, and your world is, at any given time, restricted to your immediate environment. Imagination creates the possibility of imagining the past, the future, other places. Memory, in the sense of the recall of some past event, becomes possible.

Perhaps even more important, imagination makes possible an active inner mental life, vital to consciousness. It is only because we can imagine that we are doing things, that we are not in fact doing, that inner mental action is possible at all. All thought, all consideration of possibilities (essential for science) is only possible because of the possession of the capacity to imagine.

The next crucial step in the creation of human consciousness is, I suggest, the creation of: **vi Personalistic Understanding**. This is to be understood in such a way that animals as well as humans can have personalistic understanding of each other. It might be better to speak of "animalistic" understanding. My claim is that consciousness as we humans know it comes gradually into existence with the progressive evolution of personalistic understanding, as characterized in chapter six. Personalistic understanding, in turn, is the outcome of combining *imitation* and *imagination*.

We saw, above, that imitation has survival value where there is parental care (or a social life, more generally). We saw, also, that imitation makes cultural evolution possible, thus already being of fundamental importance in our creation (in that we are above all the outcome of a long process of cultural evolution associated, especially, with the evolution of language). It deserves to be noted that imitation, even in its most elementary and purposive form, has elements of personalistic understanding. Even if I am a robot, in order to imitate another I must, in a control, purposive sense, regard the other as another *me*, use the actions of the other as a model for my actions. In imitating another, we will even be reproducing in our brains control processes functionally analogous to control processes going on in the other's brain.

With sentience and imagination, something like full-blooded personalistic understanding becomes possible. Creature A achieving personalistic understanding of creature B can be described as: A imitates B in his imagination, even to the extent of imitating what B is experiencing and imagining. When imitation is carried to the lengths, in other words, of being an imitation, in imagination, of the other's inner actions as well as outer actions, then it constitutes personalistic understanding.

It deserves to be noted that if personalistic understanding is a kind of imagined imitation, it is also not so very different from imagination *per se*. To imagine that I am doing something that I am not doing is not so very different from imagining that I am some other being, doing and experiencing what the other being is doing and experiencing.

Personalistic understanding, as characterized above as a kind of imagined imitation, makes it possible for the inner self to absorb features and characteristics of other inner selves; new features of inner selves that come into existence as a result can, in turn, evolve culturally (imitation being the key to cultural evolution). There is here a mechanism for the progressive creation of features of the inner self that go to make up human consciousness by means of a process of cultural evolution.

More specifically, I suggest, personalistic understanding creates "self-consciousness",

generally regarded as the distinguishing feature of human consciousness.

What happens, I suggest, is this. In improving our personalistic understanding of others we encounter others' personalistic understanding of ourselves. We encounter, in other words, the outer, public, other conception of ourselves. This, we will discover, differs from our own conception of ourselves. As a result of becoming aware of ourselves *as understood by others,* we become aware of ourselves *as we experience and understand ourselves*. And this latter is self-consciousness. It is in other words the *contrast* between others' awareness of ourselves and our own awareness of ourselves that enables us to be aware of *our own* awareness of ourselves - thus becoming aware of our own self-awareness. (Self-consciousness is perhaps awareness of self-awareness, rather than just self-awareness itself.) Without being able to see ourselves from the perspective of others, we are not able to notice that there is also our own perspective of ourselves, fundamentally different from all the other perspectives in that it includes self-awareness.

Many young children go through a stage of being "self-conscious", as it is described. This is acute awareness of others' perception of oneself, possibly accompanied by embarrassment, shyness or flirtatiousness. Above I have in effect argued that self-consciousness, the genuine article, is a product of something that is often called "self-consciousness" but is actually acute awareness of others' awareness of oneself.

On this view, our own self-consciousness, that which we are most inclined to identify as our unique, innermost self, nearest to being our soul, our highly personal and individual identity, is actually something profoundly *social* and *historical* in character, made possible by personalistic understanding, and the cultural evolution of self-hood which is in turn made possible by personalistic understanding.

The close linking of human consciousness, including self-consciousness, to personalistic understanding, and therefore to the *social,* has the consequence that human consciousness acquires an *historical* dimension. The kind of awareness we have of the awareness that others have of ourselves can develop, this leading to a development in our own self-consciousness, in turn requiring a deepening of personalistic understanding. The character of human consciousness grows through history. We progressively bring each other into existence (the self being understood here to be the self-conscious self).

All this is facilitated by the development of:

vii Communication and Language. Indeed, the development of personalistic understanding (and so consciousness) and the development of communication and language provoke each other.

It is important to appreciate the multi-layered character of human communication, and therefore the step-by-step stages in which human communication can be built up from primitive animal communication. Let us suppose A communicates to B. The following stages can be distinguished.

(i) A acts in its own interests, for example goes rapidly into flight to avoid a predator; B takes this behaviour as an indication of something (in this case danger), for him, and acts accordingly.

(ii) In addition, A does something which is such that the sole purpose of it is to communicate to B, even though A has no such conscious intention. Here A might squawk as it goes into flight in a manner characteristic for that species in such circumstances; B reacts accordingly.

(iii) In addition, A has the purpose of signalling to B since, if A knows that it is on its own it will not signal (e.g. squawk).

(iv) In addition, A has the purpose of communicating the message of the action to B, so that,

in the case of the squawk, the bird squawks in order to warn B. If B is present but in no danger then the bird does not squawk.
(v) B understands the message, the meaning of the squawk.
(vi) A has the purpose of B understanding the meaning of the message.
(vi) B understands this too.
(vii) A intends B to understand this.

And so on (the multi-layers of mutual understanding, initially profoundly significant becoming, as one goes on further, increasingly insignificant).

As Grice has shown,[31] human communication involves, quite essentially, multi-layers of mutually understood intentions. If I am to communicate with you by means of language, I must intend this, you must understand that I intend it, and I must understand that you understand. The progressive development of human communication through these stages from its beginnings in primitive animal communication *is* the progressive development of personalistic understanding and self-consciousness. In this sense, the progressive development of communication, first without and then with language, *is* the progressive development of self-consciousness. This progressive enrichment of communication is, at the same time, the progressive enrichment of meaning and knowledge. There is meaning and knowledge even at stage (i), but it is meaning and knowledge of a very meagre kind (without cultural evolution, and therefore not involving anything like memes); once stage (vii) is reached, and assuming that "A intends B to understand that A has the purpose of B understanding the meaning of the message" is all to be understood personalistically, A and B both being sentience and consciousness, then something close to full human meaning and knowledge have been attained.

9 Implications for Our Understanding of Our Human World Today

This concludes my quasi-Darwinian account of the evolution of sentience, consciousness, communication, language and free will in the physical universe.[32] I finish with three further points that affect our understanding of our human world today.

viii The Discovery of Death. Imagination makes it possible to become aware of death even when it is not threatened. It makes it possible to discover the eventual inevitability of death. This discovery will be immensely aided by developments in personalistic understanding and language. Given the fundamental role that survival plays in evolution, the discovery that the pursuit of survival is ultimately doomed to fail can only be traumatic. Much of the effort of human culture is devoted to denying death, transforming it into a journey, into an unimportant event. It will be both important and difficult for human consciousness to come to terms with the inevitability of death.

ix Clash of Controls. Mammals, we may presume, act as they do as a result of motivational control. There are, in effect, two control systems: the master control system of hormones, levels of such things as sugar and oxygen in the blood, which controls emotions and motivations; and the servant control system, the functioning (sentient and partially conscious) brain of the animal. Mammals plan their immediate actions, but do not plan their way of life; they do what they are prompted to do by their feelings and desires, and these come and go in such a way that the animal ends up living the way of life it needs to live in the given environment in order to survive. As long as the conscious brain does not have a vivid and stable enough imagination to be persistently aware of months into the past and future, it will not be able to be conscious of the way of life; it therefore cannot be consciously planned and controlled.

But once the conscious brain, intended by evolution to be the servant control system,

becomes aware of events stretching years into the past and future, it becomes possible for the conscious brain to attempt to plan the way of life. Once the attempt is made, the servant control system seeks to become the master control system. Conflict is inevitable. This is the source of the conflict inherent in being human.

x From the Tribe to the Modern World. Our psyches, our emotional, motivational and intellectual makeup, were designed to work well granted that we live in a hunting and gathering tribe, where everyone knows everyone else, and life proceeds from day to day. It is reasonable to assume that ancestors which we have in common with other primates today lived, millions of years ago, in this fashion. The hunting and gathering way of life is basic to our human nature.

In a multitude of ways, the modern world differs dramatically from hunting and gathering tribal life. Many of the problems of the modern world are due to the fact that we are, emotionally and motivationally, ill-adapted to living in such a world. The situation is made worse by our lack of understanding of the source of our difficulties. The evolutionary framework provided here for the understanding our ourselves has, potentially, I believe, immensely important implications for our capacity to cope with our problems, and become, globally, a little wiser.[33]

10 Science and Reason in the Physical Universe

The account, given above, of how non-sentient purposiveness evolves into sentience, personalistic understanding, consciousness, free will, communication and language assumes also, of course, that insentient *belief*, implicit in the actions of a merely purposive thing becomes, in turn, sentient belief, conscious belief, inter-personal belief, belief captured in language, belief sufficiently accurate and critically scrutinized to be deemed *knowledge*, even *science*. The evolution of human knowledge and reason from their biological, purely purposive origins is intimately bound up with the evolution of consciousness, free will, personalistic understanding and communication.

11 The Evolution of the Darwinian Mechanisms of Evolution

This chapter may be summarized as follows.

In order to understand the intricate and all but miraculous dovetailing together of (1) the physical and the purposive apparent in all life, and (2) the physical and the personalistic apparent in all sentient life, we need to appeal to Darwin's theory of evolution. But, in order to be capable of explaining the evolution of purposiveness, sentience, consciousness, personalistic understanding and free will, Darwinian theory must be reinterpreted. It needs to be interpreted as a theory that is intended, at a fundamental level, to explain the evolution of the purposive and the personalistic in the physical universe. In order to do this, Darwinian theory exploits physical, purposive and personalistic modes of explanation. However, if the Darwinian pattern of explanation for evolution uses purposive or personalistic explanations, at any evolutionary stage, these purposive or personalistic explanations must be couched in terms of purposive or personalistic phenomena whose evolution has already been given a Darwinian explanation. (Darwinian theory must not presuppose what it seeks to explain.)

As evolution proceeds, and living things become more and more diversely and richly purposive and personalistic, the Darwinian mechanisms responsible for evolution themselves evolve. They cease to be wholly blind and mechanistic, and gradually incorporate elements of the purposive and personalistic. Above, in section 6, I indicated how this evolution of evolutionary processes comes about. I conclude this chapter with a summary of some of the main steps involved in the gradual, evolutionary incorporation of the purposive and personalistic into the mechanisms of evolution.

The Darwinian mechanisms for evolution consist, essentially, of two parts: (a) mechanisms of *reproduction*, including reproduction of persisting, inheritable variations; and (b) mechanisms of *selection*. Both evolve with evolution. When evolution first got going, nothing as complex as the DNA molecule, plus attendant molecular machinery involved in reproduction, can have existed. Initially mechanisms of reproduction involved very much simpler molecular processes. Genes, DNA, RNA, ribosomes and the genetic code (all a part of the mechanism of reproduction today) are the outcome of a long process of early evolution, perhaps for something like a billion years, that has left almost no trace. Subsequently, such things as the evolution of multi-celled creatures (posing new problems of reproduction), sex, and the gestation of the fertilized egg in the womb, all amount to modifications of the mechanisms of reproduction. Here, however, we are concerned with those steps in the evolution of the mechanisms of evolution that involve the gradual increase in the involvement of the purposive and personalistic in these mechanisms.

It is vital to appreciate that to say that the purposive is *involved* in the mechanisms responsible for evolution is not to say that these mechanisms are *fully* purposive in character. As we shall see, very many evolutionary steps need to take place, each slightly increasing the involvement of the purposive, before the blind, wholly purposeless mechanisms of evolution, that are in operation immediately after the origin of life, are transformed into fully purposive mechanisms of evolution.

Some important steps in the gradual incorporation of the purposive and personalistic into the Darwinian mechanisms responsible for evolution are the following.

1. Blind reproduction (with inheritable variations) and natural selection: initially wholly purposeless, the molecular processes involved in reproduction initially very different from, and much simpler than, those in operation today.
2. Blind reproduction (with inheritable variations) and natural selection, the latter having an element of purposiveness associated with it, due to the fact that what has survival value may depend on what purposes are being pursued, so that a persisting change in purposeful activity may contribute to subsequent inherited bodily changes (as in the cartoon example, discussed above in section 6, of the dog-like creature becoming a beaver). *Changes* in purposeful activity that persist across generations may come about because of (a) genetic changes, (b) changes in the environment, (c) changes made to the environment by the living things in question (d) movement to a new environment (e) changes in purposeful activity due to individual learning, and imitation. Subsequent inherited variations (the result of blind, random, purposeless occurrences) may only proliferate because of a prior change in purposeful activity, which has itself come about in one or other of ways (a) to (e). In this case, the bodily change is due, in part, to the prior change in purposeful activity. Purposefulness contributes to the evolutionary change, even though the mechanism for the evolutionary change overall is, of course, not purposeful.[34] In the case of (a), the contribution of purposefulness is at a minimum, because the change in purposeful activity occurs in a purely blind, random way. As one goes from case (b) to (c), (d) and (e), the contribution of purposefulness is gradually increased. The role of purposeful explanation in the Darwinian explanation for the evolutionary change is slightly increased.
3. As above, but the processes of selection involve, not just purposeless physical phenomena plus the purposeful actions of the evolving living thing in question (as in 1 and 2) but, in addition, the actions of other living things, such as predators, prey, potential mates, parents. The result, as we saw in section 6 above, is unknowing breeding. Sometimes the outcome of breeding is in the interests of the breeder, as in the case of parents selecting offspring, and

sexual selection. More often, perhaps, the outcome is against the interests of the breeder, as in the case of the predator/prey relationship.

4. As above, but reproduction involves parental care. Initially, parental care may merely increase the chances of survival of offspring, as in the case of crocodiles perhaps. But it then becomes essential, as in the case of birds and mammals. Parental care may be under genetic control, but still involves purposive action. The element of purposiveness increases, however, when parents learn how to exercise good parental care, and when parental care depends on the parent having been a recipient of parental care when young. (In the latter case, an offspring reared without parents becomes incapable of parental care itself.)

5. As above, but reproduction (of way of life) also involves, in childhood especially, *imitation* of purposeful activity of parents or other adults, this activity the outcome of individual learning of some ancestor, or series of ancestors. This step involves a dramatic injection of the purposiveness into the mechanisms of reproduction, and hence into the mechanisms of Darwinian evolution. This new aspect of reproduction, based on learning and imitation, may be called *cultural*. Its incorporation into the mechanisms of evolution has a dramatic effect on evolution (as I indicated in section 6 above) because it introduces a quasi-Lamarckian character to evolution. An animal can learn a new trick (catching termites by putting sticks into termite nests) which is then passed on to offspring by means imitation: an acquired characteristic is inherited (culturally).

As I remarked above, the instinct to imitate is likely to have survival value where there is parental care (and in any case requires at least that much social life). Once there is parental care, in other words, the evolution of the instinct to imitate, and hence the possibility of cultural reproduction and evolution, can be explained along straightforward Darwinian lines. Thus, the evolution of parental care leads to the evolution of cultural reproduction and evolution, and the quasi-Lamarckian kind of evolution that results.

It may be objected that it is very artificial - little better than a play on words - to put "cultural reproduction" on a par with ordinary "biological" reproduction", the result of sex, fertilization of eggs, gestation, birth and growth to maturity, all under genetic control. This objection is likely to be made by those who see evolution as primarily evolution of bodily structure, or of genes, DNA molecules and other molecular structures, physiology, design. Construing what evolves in such "biological hardware" terms leads one, when it comes to reproduction, to concentrate on processes that reproduce such "biological hardware", namely the conventional processes of biological reproduction.

If one adopts this "hardware" attitude towards what evolves then, when one is confronted by cultural evolution (evolution by cultural means), the response is likely to be that entities of a novel kind have begun to evolve, cultural entities which Dawkins has dubbed "memes".[35]

It is, however, not just legitimate, but of fundamental importance (as I have argued above), to interpret neo-Darwinism as a theory which explains how and why *purposiveness* has evolved in nature. Darwinian theory gets off the ground, and only gets off the ground, when entities come into existence which can be construed to be, more or less successfully, pursuing the goals of survival and reproductive success. Darwinian theory, and biology more generally, is about *life*. In considering reproduction, within the context of Darwinian theory, what matters is the reproduction of *ways of living*, modes of goal-pursuing in the world. Granted this perspective, whatever is essential to the reproduction of a *way of life* must be judged to be an essential component of the Darwinian notion of "reproduction", and thus an essential part of the Darwinian mechanism of evolution (at a given stage in its evolution). Becoming an adult chimpanzee is not only a question of conception, embryonic development,

birth and bodily growth; it also involves quite essentially learning how to act as an adult chimpanzee. Without cultural inheritance (childhood learning through imitating adult chimpanzees) this would not be possible. Being able to survive and reproduce may well depend on cultural inheritance: bereft of the opportunity of learning from adult chimpanzees through imitation, a chimpanzee may well be incapable of surviving and reproducing in the natural environment.

Evolutionary changes that come about solely as a result of cultural evolution may play an essential role in subsequent evolutionary changes due to mutations. (Variations only have survival value because of earlier changes in the way of life due to cultural evolution.) I gave, above, a fictional example of this effect: the dog-like creature becoming a beaver. It is reasonable to suppose, as I suggested above, that this interplay between cultural evolution, to be understood in purposive terms, and subsequent gene-based evolution, played a very important role in the evolution of the human capacity to learn, and develop, language. What distinguishes us from all other living things is that we are the product of a massive process of cultural evolution, linked in the way I have just indicated to gene-based evolution. Human consciousness, in so far as it differs from chimpanzee consciousness, is made possible by the evolution of language. We only exist as human beings, in other words, because of the evolution of cultural evolution.

6. Steps 2 to 5 involve increasing incorporation of purposiveness into the Darwinian mechanisms responsible for evolution. At some stage, however, feeling and desire enter the arena: the merely purposive becomes personalistic. Steps 2 to 5 become further steps in the evolution of the mechanisms of evolution as the merely purposive is replaced by the personalistic. And as such things as imagination, language, personalistic understanding and consciousness emerge, the kind of actions, open to being understood personalistically, that can be incorporated into Darwinian mechanisms of evolution, become more and more richly personalistic. Thus breeding, which is, to begin with, unintended and purely purposive, becomes, with conscious human beings with some knowledge of breeding methods and results, a wholly conscious, intended process.

7. Once language and human culture are in place, massively facilitating the quasi-Lamarckian processes of cultural evolution, it becomes possible for people to engage in projects that make progress across many generations. Natural science is one strikingly successful example.

8. Genetic engineering, made possible by scientific progress, in turn a result of cultural evolution, renders the process of gene modification the outcome of conscious (personalistic) action.

9. Not all personalistic modifications of Darwinian mechanisms of evolution that are conceivable, have actually occurred. One can imagine, for example, that humanity as a whole may one day discover how to take control of its own cultural evolution, its future, in such a way that progress is made towards global wisdom, civilization and enlightenment. How this might come about will be discussed in the next but one chapter.

References

Boyd, R. and Richerson, P. J. (1985) *Culture and the Evolutionary Process*, The University of Chicago Press, Chicago.

Cairns-Smith, A. G. (1986) *Seven Clues to the Origin of Life*, Cambridge University Press, Cambridge.

Calow, P. (1983) *Evolutionary Principles*, Blackie, Glasgow.

Chalmers, D. (1996) *The Conscious Mind*, Oxford University Press, Oxford.

Clark, A. (1997) *Being There*, 1997, MIT, Cambridge, Mass.
Darwin, C. (1976) *The Origin of Species*, Penguin, Harmondsworth (first published 1859).
Dawkins, R. (1978) *The Selfish Gene*, Paladin, London.
_____ (1986) *The Blind Watchmaker*, Longman, Harlow, Essex.
Deacon, T. (1998) *The Symbolic Species*, Penguin, London.
Dennett, D. (1991) *Consciousness Explained*, Allen Lane, London.
_____ (1996) *Darwin's Dangerous Idea*, Penguin, London.
Dunbar, R. (1997) *Grooming, Gossip, and the Evolution of Language*, Faber and Faber, London.
Eigen, M. (1992) *Steps Towards Life*, 1992, Oxford University Press, Oxford.
Grice, H. P. (1957) 'Meaning', *Philosophical Review 66*, 377-388.
_____ (1989) *Studies in the Way of Words*, Harvard University Press, Cambridge, Massachusetts.
Hardy, A. (1965) *The Living Stream*, Collins, London.
Holdcroft, D. and Lewis, H. (2000) 'Memes, Minds and Evolution', *Philosophy 75*, 161-182.
Kohler, W. (1927) *The Mentality of Apes*, Kegan Paul, London.
Küppers, B. (1990) *Information and the Origin of Life*, 1990, MIT Press, Cambridge, Mass.
Lipson, H. and Pollack J. B. (2000) 'Automatic design and manufacture of robotic lifeforms', *Nature 406*, 974-978.
Maxwell, N. (1984) *From Knowledge to Wisdom: A Revolution in the Aims and Methods of Science*, Basil Blackwell, Oxford.
_____ (1985) 'Methodological Problems of Neuroscience', in *Models of the Visual Cortex*, ed. D. Rose and V.G. Dobson, John Wiley, Chichester, 11-21.
Maynard Smith, J. (1975) *The Theory of Evolution*, Penguin, Harmondsworth.
Maynard Smith J. and Szathmary, E. (1995) *The Major Transitions in Evolution*, W. H. Freeman, Oxford.
Mayr, E. (1982) *The Growth of Biological Thought*, Harvard University Press, Cambridge, Massachusetts.
Monod, J. (1974) *Chance and Necessity*, 1974, Fontana, Glasgow.
Pinker, S. (1998) *How the Mind Works*, Penguin, London.
Popper, K. R. (1972) *Objective Knowledge*, Oxford University Press, Oxford.
_____ (1977) Part I of K. Popper and J. Eccles, *The Self and Its Brain*, Springer-Verlag, Berlin.
Skinner, B. F. (1973) *Beyond Freedom and Dignity*, 1973, Penguin, Harmondsworth.
Wooldridge, D. E. (1963) *The Machinery of the Brain*, McGraw-Hill, New York.

Notes

1. For free action, the following must be the case. Whenever we decide to perform some action A and as a result perform it, the brain processes that are the decision to do A must control muscles of the body so that action A results. Whenever we ponder whether to do A or not, explore reasons for and against doing A, and come to a rational decision, the sequence of brain processes that are these processes of rational pondering and decision-making must occur in just the right way to constitute genuinely rational pondering and decision-making. And vastly more demanding, for free action, the personalistic and physical must be matched so intricately, extensively and delicately that all (or almost all) counterfactuals corresponding to the personalistic explanation of the free action are included in the counterfactuals of the physical explanation of the action. It is not just what actually happens that must match up precisely, as

far as the two kinds of explanation are concerned; what would have happened if such and such had been the case, for infinitely many counterfactual, possible states of affairs, must match up precisely as well.

2. The two-aspect, or multi-aspect solution to the human world/physical universe problem that I am expounding in this book is, I claim, an immense improvement over Popper's "three worlds" view, which can be regarded as an attempt at solving the same problem, and is expounded in such works as: Popper (1972), (1977). Popper's three worlds view is committed to the highly implausible idea that there exists a quasi-Platonic world of ideas, arguments, problems, etc., which exists largely independently of the material world of mind. As we saw in chapter five, horrendous problems are created by this view as to how Darwinian evolution can possibly have created this new, almost autonomous world of ideas which, once created, actually *interacts*, with the physical world via physical processes occurring in conscious brains. None of these difficulties faces the duo comprehensibility, compatibilist, experiential physicalist view that I am expounding and defending here.

3. In chapter five I explored, and rejected, the possibility of extending physical and personalistic explanations so as to explain correlations between physical and mental aspects of brain processes.

4. For Dennett (1996), Darwin's dangerous idea is a kind of "universal acid" which threatens to transform, even destroy, everything that we hold dear. The perspective of this book is rather different. It is physicalism, the idea of Smart, Laplace, de la Mettrie or, to trace it back to its origins, Democritus, that is dangerous, threatening to reduce our human world to mere physics, thus annihilating all meaning and value. Darwin's wonderful idea explains how purpose, meaning and value have gradually emerged or evolved in a physicalist universe initially devoid of purpose, meaning and value. (Dennett's brilliant book does, perhaps, on its last page, come to a similar conclusion.)

5. The second and third of the three arguments for holding that personalistic explanations cannot be reduced to physical explanations, indicated in section 6 of the last chapter, establish also, when appropriately modified, that purposive explanations cannot be reduced to physical explanations. Chalmers has argued very persuasively that the biological supervenes logically on the physical: see Chalmers (1996), 35 and 73. If this is correct, does not this mean that biological and purposive explanations are reducible to physical explanations? No. In asserting that the biological supervenes logically and globally on the physical one is asserting that it is inconceivable that *all* the physical facts of this world remain the same but some biological facts are different. But explanations, whether physical or purposive, refer not just to the facts of this world, but to a range of counterfactual situations, to a range of possible worlds. The two arguments indicated in the last chapter exploit this feature of explanation to establish that purposive (and, to that extent, biological) explanation is not reducible to physical explanation.

It would seem reasonable to conclude from this that logical supervenience is a necessary, but not a sufficient condition for reducibility of explanation. This is what Chalmers concludes: see Chalmers (1996), 48. But this may be doubted. If one gives purposive and biological explanations the kind of conjectural essentialistic interpretation indicated in ch. 3 in connection with physical explanation, then purposive and biological facts carry implications about counterfactual situations. In this case, whether or not the purposive and biological supervene logically on the physical might depend on what counterfactuals are taken to be implicit in purposive and biological facts. The thesis that the purposive and biological supervene logically

on the physical is perhaps not as obvious as Chalmers supposes. (Throughout this note I assume that the purposive and biological are characterized in a way that is free of perceptual or sensory qualities, and sentience.)
6. More will be said in support of this claim below.
7. We may define living beings to be naturally occurring purposive beings. Or they may be defined to be purposive beings able to reproduce themselves, which would make all current artificial purposive devices, such as robots, non-living, but which would make it possible for artificial life, in the defined sense, to be created in the future. A step in this direction has been taken by Lipson and Pollack (2000), who have created robots able, in some respects, "to sustain their own evolution".
8. When the programme of artificial intelligence began, in the 1950's, the basic idea was understood to be that of the feedback mechanism, which explained how purposiveness could exist in a way that is compatible with physicalism. The AI movement then got sidetracked into *computationalism*, probably because of commercial pressures to design computers that functioned in response to programmes. The original insight was lost sight of, and the vital link with life and evolution fell into decay. Even as intelligent an author as Steven Pinker defends the computational theory of mind: see Pinker (1998), ch. 2. The very name of the discipline "artificial intelligence" is a misnomer - I hesitate to say an indication of lack of intelligence. The discipline ought to have been called, from the outset, artificial life, or artificial control. See Maxwell (1985): and see Clark (1997).
9. For accounts of Darwinian theory, see: Darwin (1976), Maynard Smith (1975), Mayr (1982), Calow (1983), Dawkins (1978).
10. For a delightful, if unorthodox, speculative theory about the origin of life see Cairns-Smith (1986). See also Eigen (1992) and Kuppers (1990).
11. It is striking that nowhere in his brilliant *The Blind Watchmaker* (1986) does Richard Dawkins attribute purposiveness to living things. Throughout the book, the problem that Darwin's theory is held to solve is the problem of *design*. But this is an inadequate way of construing Darwin's theory. The very concept of design presupposes purpose: it is only when we know what the purpose of something is, that we can assess whether or not it is well-designed. Furthermore, in a physicalist universe, things can only pursue goals (be purposive) if they are more or less intricately designed to do so (embodying feed-back mechanisms, sensing devices, and the capacity to act or grow, at least to the extent of throwing a switch as in the case of the thermostat). Finally, as I have just argued in the main text, Darwinian theory only makes sense when applied to things that are at least construed to be goal-pursuing, in that it requires these things to be interpreted as having the fundamental goal of reproductive success.
12. Darwin did not himself see his fundamental problem in quite this light. For Darwin, the basic problem was the origin of species, as enshrined in the very title *The Origin of Species* (Darwin (1968), first published in 1859). Darwin understood, however, that his theory explained the manner in which living things are so beautifully and diversely adapted to their diverse environments and ways of life. Adaptation presupposes purpose.
13. The greatest defect in Darwin's presentation of his theory lay in his ignorance of the laws of inheritance, discovered by Mendel and published in an obscure Austrian journal in 1865, but generally ignored until the 20th century. Lacking knowledge of the Mendelian rules of inheritance, Darwin had no adequate reply to the criticism that new variations would tend to disappear across generations as they blended in with pre-existing variations. In later editions of

his *Origin of Species*, Darwin even introduced elements of Lamarckism into his account of his theory, in an attempt to meet this and other criticisms of the original version of his theory.
14. Darwinian theory is no doubt falsifiable. Observing inheritance of acquired characteristics would refute the theory, in that it would refute the claim that the Darwinian mechanisms responsible for evolution are the *only* evolutionary mechanisms in operation. Nevertheless, Darwinian theory is not predictive in the way that a dynamical theory of physics, such as Newtonian theory, is predictive, in that it predicts how a system, such as the solar system, evolves in time. Darwinian theory does not predict the course that evolution will take. Furthermore, Darwinian theory is parasitic upon other descriptive and explanatory accounts of phenomena, physical and purposive, and does not issue in its own descriptions of phenomena, as fundamental physical theories do.
15. It is legitimate, I think, to attribute this "purposeless" view to Monod (1974) and Dawkins (1986). Dawkins, in his defence, might claim that he wishes only to argue that the mechanisms of evolution are "blind" and purposeless, not the products, namely living things. This reply is fair enough. Nevertheless, as I have already indicated, Dawkins fails to emphasize the vital point that genuinely *purposive* beings are generated as a result of the operations of the purposeless mechanisms of evolution. Dawkins interprets Darwinian theory as solving the problem of design in the biological world, and the problem of why living things should be so well adapted to their environment. Both design and adaptation, in this context, presuppose that living things have the life-goals of reproductive success. But Dawkins does not make this point. Nor does he assert, much less emphasize, that the fundamental problem addressed by Darwinian theory is: How and why have the diverse purposive beings, that living things are, evolved in an ultimately purposeless universe?
16. This is similar to, but not precisely the same as, the Baldwin effect, discussed by Dennett (1996), 77-80. Dennett does not take into account that the dog-like creature may learn to catch fish from its parents by imitation.
17. I am here, in the interests of clarity, contracting to a two-step change what would, in reality, be a many-step change, gradual changes in behaviour across generations being interwoven with a number of small changes in physiology being brought about by a number of mutations in successive generations: the basic point remains valid in the more complicated, realistic process of evolution.
18. For a marvellous account of the way in which learnt behaviour can affect subsequent evolution, along the lines indicated here, see Hardy (1965).
19. Darwin begins the great argument for his theory, in *The Origin of Species*, with a discussion of *our* breeding of animals. (The chapter is called "Variation under Domestication".) Breeding has sometimes been interpreted as amounting to no more than a sort of metaphor for natural selection: what we do, consciously, namely select those offspring with characteristics we prize, Nature does "blindly", in selecting those offspring best able to survive and reproduce. But what I am arguing is that there is much more to it than this; unintended, unconscious breeding of living things by living things is an all-pervasive feature of evolution. It is one which we can only do proper justice to if we acknowledge the role that purposive explanation plays in evolution and biology - where purposiveness is understood in its compatibilist, non-Aristotelian sense. When we intentionally and consciously breed animals we are modifying and developing an activity that is a long-standing, pervasive and highly significant feature of evolution, namely the tendency of *all* living things, *un*intentionally and *un*consciously, to breed both their own

species and other species. The (historical) explanation for our existence as purposive beings must include this tendency of life to breed itself (unintentionally) into existence.
20. Dawkins, famously, has argued that "the fundamental unit of selection" is not "the individual" but "the gene, the unit of heredity" (Dawkins, 1978, 12). One might perhaps say that Dawkins' view is that what is selected for are genes good at reproducing themselves, even though selection acts on living things produced, in part, by genes. But however one decides the debate as to whether the unit of selection is the gene or the individual, one thing is clear: the fundamental task of Darwinian theory is to explain the existence and evolution of living things, purposive life (and not genes). It would be peculiar indeed to hold that what fundamentally needs to be explained is the existence in the world of long strings of DNA molecules - explaining the characteristics of living things being but a means to that end. "The unit of selection" should not be conflated with "what Darwinian theory is fundamentally about and seeks to help explain".
21. None of these disanalogies is quite conclusive. Thus the distinction between genome and phenotype would not arise for the earliest forms of life; lines of descent, having separated, do sometimes subsequently merge in the biological world, as the existence of mitochondria in our cells indicate. (Mitochondria, at one stage distinct life forms were, at some stage, absorbed into the ancestors of our cells.) From the perspective of this book, the real objection to seeing memes as entities that undergo Darwinian evolution is that memes are not purposive things, and Darwinian theory is restricted to things that are purposive. But this objection presupposes what I am trying to establish.
22. What Holdcroft and Lewis are criticizing, of course, is not any account of the evolution of culture, but any *Darwinian* account which takes elements of culture as the things that undergo Darwinian evolution. (To this I would add that any history of culture should always be pursued as the history of an aspect of life, life being what is fundamental, and culture being an aspect of (some) life: see note 23.)
23. In ch. 9, and at greater length in Maxwell (1984), I argue for a kind of inquiry that gives intellectual priority to life and the problems of living, and has, as its basic aim, to promote wisdom. From this perspective, the superiority of interpreting Darwinism as being, fundamentally, about evolving purposive (and eventually personalistic) life is at once apparent.
24. That "cultural evolution" is quasi-Lamarckian in character has often been pointed out: see, for example, Boyd and Richerson (1985, 8) and Dennett (1991, 355). But these authors are making a point rather different from the one I make when I assert that cultural evolution is Lamarckian in character, because for these authors "cultural evolution" means "the evolution of culture" and not, as for me, "evolution by cultural plus genetic mechanisms". Dennett makes this point quite explicitly as follows: "Usually the 'charge' that cultural evolution is Lamarckian is a deep confusion ... [but when] it is undeniable ... the entity that exhibits the Lamarckian talent of passing on an acquired characteristic is not the human agent, but the meme itself" (Dennett, 1996, 355, n. 6).
25. For excellent accounts of the origins and evolution of language see Dunbar (1997) and Deacon (1998).
26. It is unlikely that this process (of cultural evolution creating the circumstances necessary for subsequent mutations to lead to genetic evolution) is confined to the evolution of human beings. It can occur whenever animals are able to learn individually, and learn by imitation. The (hypothetical) example, given above, of a dog-like creature evolving into a beaver-like creature

as a result, in part, of the prior discovery that fish can be caught and are good to eat, is an example of cultural evolution making possible subsequent genetic evolution. The dog-like creatures learn to catch fish as a result of individual learning and imitation: this change in the way of life is pure cultural evolution. It is, however, a necessary first step in the subsequent evolution into the beaver-like creature: flipper-producing mutations would not lead to beaver-like creatures without the prior occurrence of the cultural evolution towards fish-catching dog-like creatures.

27. For a rather different account of various stages of evolution, see: Maynard Smith and Szathmary (1995).

28. See Wooldridge (1963), 82-83.

29. Descartes' "I think, therefore I am" needs to be replaced by Tolstoy's "I desire, therefore I am".

30. See Kohler (1927).

31. Grice (1957), reprinted in Grice (1989), ch. 14. Grice makes no attempt, however, to indicate, as I have done here, the manner in which the multi-layered character of human communication can be seen as having emerged gradually as a result of Darwinian evolution.

32. The speculative account of the evolution of sentience, consciousness and free will just indicated is a development of the account first given by me in Maxwell (1984), 174-181 and 267-275: see also Maxwell (1985). What I have done is to show how it is possible for neo-Darwinism, when appropriately reinterpreted, to explain the evolution of purposiveness, sentience, consciousness, meaning, personalistic understanding, and free will.

33. This theme is developed further in Maxwell (1984), and in chapter 9 of the present book. See also Pinker (1998), ch. 6-8.

34. As I have already remarked, "Involving purposive activity" must be distinguished from "being purposeful". That the mechanism of evolution "involves purposive activity" (the explanation for an evolutionary change exploiting some purposive explanation) does not mean at all that the mechanism in question is, overall, purposive in character. "Involving purposive activity" comes in finely graded degrees; it is only when many of these degrees are added together that one has a mechanism of evolutionary change that is fully purposive in character.

35. Dawkins (1978), 206-215.