The anthropic principle and its implications for biological evolution

By B. Carter, F.R.S.

*Groupe d'Astrophysique Relativiste, Obervatoire de Paris – Meudon,
5 Place Jules Janssen, 92 Meudon, France*

In the form in which it was originally expounded, the *anthropic principle* was presented as a warning to astrophysical and cosmological theorists of the risk of error in the interpretation of astronomical and cosmological information unless due account is taken of the biological restraints under which the information was acquired. However, the converse message is also valid: biological theorists also run the risk of error in the interpretation of the evolutionary record unless they take due heed of the astrophysical restraints under which evolution took place. After an introductory discussion of the ordinary ('weak') anthropic principle and of its more contestable ('strong') analogue, a new application of the former to the problem of the evolution of terrestrial life is presented. It is shown that the evidence suggests that the evolutionary chain included at least one but probably not more than two links that were highly improbable (*a priori*) in the available time interval.

1. Introduction

A key event in the birth of modern science at the time of the renaissance was the Copernican revolution that transformed our understanding of our planetary system by de-throning the Earth from its central role in favour of the Sun. This was the beginning of a consistent effort by scientifically minded thinkers to break away from the anthropocentric prejudices that had dominated the mediaeval outlook. At the outset, this trend was entirely justified by the goal of scientific objectivity, but it soon came to be carried unduly far as people came to the point of advocating the opposite extreme point of view, consisting in the assumption that our own situation in the Universe is not in any way privileged, but is typically representative in a Universe that is entirely homogeneous apart from minor local fluctuations. This extreme antithesis of the anthropocentric outlook was most dangerous as a source of biased thinking when it was adopted subconsciously. However, it became easier to cope with after having been formulated explicitly as the 'perfect cosmological principle' by Bondi & Gold (1948), who used it as a hypothesis in setting up the 'steady state theory'. (After having been developed in more detail by other workers, starting with Hoyle (1949), the steady state idea fell into general disfavour for a number of theoretical and observational reasons, but a more sophisticated, albeit circumscribed, version known as the 'inflationary universe' has recently resuscitated this perennially beguiling concept (Hawking 1982).)

It was in an attempt to draw attention to the need for a more balanced intermediate attitude, between primitive anthropocentrism and its equally unjustifiable antithesis that I came to introduce the term *anthropic principle* (Carter 1974) to express the notion that 'although our situation is not necessarily central it is necessarily privileged to some extent', in so much as special conditions are necessary for our very existence. The practical scientific utility of this principle arises from its almost tautological corollary to the effect that in making general inferences from what we observe in the Universe, we must allow for the fact that our observations are inevitably biased by selection effects arising from the restriction that our situation should satisfy the conditions

that are necessary *a priori*, for our existence. The term *self-selection principle* would be an alternative and perhaps more appropriate description for this hardly questionable but easily overlooked precept. (If I had guessed that the term 'anthropic principle' would come to be so widely adopted I would have been more careful in my original choice of words. The imperfection of this now standard terminology is that it conveys the suggestion that the principle applies only to mankind. However, although this is indeed the case as far as we can apply it ourselves, it remains true that the same self-selection principle would be applicable by any extraterrestrial civilization that may exist.)

In a typical application of the anthropic (self-selection) principle, one is engaged in a scientific discrimination process of the usual kind in which one wishes to compare the plausibility of a set of alternative hypotheses, $H(T_i)$, say, to the effect that respectively one or other of a corresponding set of theories $T_1, T_2, \ldots$ is valid for some particular application in the light of some observational or experimental evidence, E, say. Such a situation can be analysed in a traditional Bayesian framework by attributing *a priori* and *a posteriori* plausibility values (i.e. formal probability measures), denoted by p_E and p_S, say, to each hypothesis respectively before and after the evidence E is taken into account, so that for any particular result X one has

$$p_E(X) = p_S(X/E), \tag{1.1}$$

the standard symbol / indicating conditionality. According to the usual Bayesian formula, the relative plausibility of any two theories A and B, say, is modified by a factor equal to the ratio of the corresponding conditional *a priori* probabilities $p_S(E/A)$ and $p_S(E/B)$ for the occurrence of the result E in the theories, i.e.

$$\frac{p_E(A)}{p_E(B)} = \frac{p_S(E/A)}{p_S(E/B)} \frac{p_S(A)}{p_S(B)}. \tag{1.2}$$

Now in the practical application of this formula it is important to bear in mind that the result will not be valid unless all relevant effects of experimental bias and observational selection have been taken into account in the interpretation of the probabilities on the right hand side. In other words one must be careful to distinguish between the appropriately renormalized *a priori* probabilities that we have denoted by p_S (S for selected or subjective) which are effective here, and the raw *ab initio* probabilities, which could conveniently be denoted by p_O (O for original or objective), that one might have derived directly from the purely abstract theory without taking account of the practical details of its concrete application. The relation between the *a priori* (selected) and *ab initio* (original) probabilities of a result X is expressible, analogously to (1.1) as

$$p_S(X) = p_O(X/S), \tag{1.3}$$

where S denotes the totality of all the selection conditions that are implied by the hypothesis of application of the theory to a concrete experimental or observational situation, but which are not necessarily included in the abstract theory on which the calculation of the *ab initio* probabilities is based. The distinction on which I have been rather laboriously insisting is entirely familiar to all working empirical scientists (even though it is easily forgotten by pure theoreticians who prefer to work exclusively at the *ab initio* level rather than at the practically relevant *a priori* level). The only new element brought in by the anthropic principle is the reminder that the set of subjective selection conditions, S, should include not only the usual allowance for the limitations of our (artificial) measuring instruments but also allowance for our own limitations as living organisms.

In order to illustrate the rather abstract considerations that have just been summarized, let us start with a quite trivial example in which the anthropic aspect as such is not involved. In an

investigation of loss of wheat from a barn, let us suppose that we wish to choose between theory A, which attributes a major share of responsibility to mice, and theory B, which attributes the responsibility almost exclusively to rats, and that these theories seem equally plausible *a priori*. Imagine that we set out to acquire experimental evidence E by setting a trap, and that the first animal we catch is a mouse. Proceeding as pure theoreticians at the *ab initio* level, we would obviously assume that the likelihood of a mouse turning up first was high in the first theory, $p_0(E/A) \approx 1$, but that it would be much lower in the second theory, $p_0(E/B) \ll 1$. If we incautiously used the corresponding ratio, $p_0(E/A)/p_0(E/B) \gg 1$ in the Bayesian formula for calculating the *a posteriori* plausibility ratio from the *a priori* plausibility ratio, we would end by concluding that the mouse theory, A, was the most likely. However, an experienced empirical investigator would certainly consider the possibility of limitations in his equipment before coming to a final conclusion. It might be that all we had at our disposal was an ordinary mouse trap, so that the effective 'capture cross section' for a rat would be negligible. Taking the consequent selection conditions into account we would obtain not only $p_{\bar{s}}(E/A) \approx 1$, but also (unless mice were very rare indeed) $p_{\bar{s}}(E/B) \approx 1$. Substitution of the resulting ratio $p_S(E/A)/p_S(E/B) \approx 1$ in the formula (1.2) would leave the *a posteriori* plausibility ratio unchanged from its *a priori* value, i.e. our experiment would have failed to discriminate.

As a second illustration, I shall now describe an equally simple but non-trivial case, which constitutes what is in fact the classic example of an argument based on the anthropic principle. This example, which is logically analogous to the preceding one, is concerned with the important question (which will be dealt with again in a more conclusive manner later on) of discriminating between hypothesis A to the effect that the development of life is of common occurrence on 'habitable' planets, comparable with our own, and hypothesis B to the effect that, on the contrary, life is very rare, even in geophysically favourable conditions. The evidence, E, consists of the fact that on the only obviously 'habitable' planet we have yet been able to observe, namely our own, life does indeed exist. If future astronomical progress should one day enable us to observe a second example of occurrence of life on a randomly chosen 'habitable planet' belonging to a not too distant star in our Galaxy, the corresponding *ab initio* probability ratio, $(p_0(E/A)/(p_0(E/B)) \gg 1$, would justify the induction that hypothesis A (that life is common) was the most likely. However, so long as the only example at our disposal is our own, no such inference is permissible, since the anthropic selection principle ensures, as a virtual tautology, that one of the *a priori* conditions, S, that must be satisfied by the first planet available for investigation by us must be the prior occurrence of life, namely our own. Thus as in the previous example we obtain not only $p_S(E/A) = 1$ but also $p_S(B) = 1$, so that our observation has no discriminating power at all, and both alternatives A and B remain equally viable. (We shall show later on that when further evidence and more detailed anthropic selection effects are taken into account it is possible to infer that hypothesis A (that life is common) is actually far less plausible than hypothesis B (that life is rare), at least so long as one has no deep theoretical reason for preferring one hypothesis to the other *a priori*.) This example was discussed at length not long ago by Crick (1981) who at the time was apparently well acquainted with the concept, though not the name, of the anthropic principle. With reference to the commonly made but unjustifiable inference that A is most likely, Crick appropriately commented, 'this argument is false', adding, 'I do not know whether such a line of reasoning has a name, but it might be called the *statistical fallacy*'. A more explicit description is in fact already available, namely *failure to respect the anthropic principle*.

As a third illustration, I would mention the case that first drew my own attention to this matter, which was an equally (albeit, at the time, less obviously) unjustifiable inference by Dirac (1937) in favour of a cosmological hypothesis A to the effect that the strength of gravitational coupling diminishes with time. In this example the role of hypothesis B is taken by the orthodox assumption that ordinary general relativistic theory, with a fixed gravitational coupling strength, is appropriate. Dirac's evidence E was effectively equivalent to the observation of a remarkable coincidence which is that the order of magnitude, *ca.* 10^{58}, of the Hubble time t_H (which can be roughly interpreted as indicating the order of the age of the Universe) obtained from the expansion rate of the Universe, as measured in fundamental Planck units ($G = c = h = 1$), agrees with the $\frac{3}{2}$ power of the gravitational coupling constant specified by the square of the proton mass in the same units, i.e. $m_p^2 \approx 10^{38}$. According to Dirac's hypothesis, A, this gravitational coupling varies so as to preserve this relation with the Hubble time as the latter increases with the age of the Universe, thereby giving the result, $p_0(E/A) \approx 1$, that was desired. Whereas in the orthodox theory B the changing Hubble age can agree with the fixed power of the coupling constant only at one particular epoch so that the *ab initio* likelihood of the occurrence of the coincidence is low, i.e. $p_0(E/B) \ll 1$. However, as in the preceding examples, this does not justify discrimination against B in practice, because there is a selection condition (of an anthropic nature) which ensures that the relevant *a priori* probability is nevertheless of order unity, i.e. $p_S(E/B) \approx 1$ as before. The selection effect in question, which was first pointed out by Dicke (1961) consists in the condition that biological systems based on the same principles as our own can hardly come into existence except when the age of the Universe is of the same order as the well known main-sequence (hydrogen burning) lifetime, $\tau_0 \approx m_p^{-3}$, of a typical ordinary star. (The reason, briefly, is that some stars must already have burned out to provide the medium elements that are essential for our chemical constitution, while the maintenance of a suitable continuous energy supply requires that some stars – including the Sun in our own case – must still be burning.)

2. THE STRONG ANTHROPIC PRINCIPLE: A DIGRESSION

With reference to the analysis that has just been presented of the significance of Dirac's large number coincidence, it is often objected that Dicke's restriction on the cosmological epoch is only applicable to life forms based on principles similar to our own, and need not be valid for quite different life forms whose base of support might be quite different from an ordinary star–planet system. It must certainly be admitted that the difficulty of imagining the technical mechanisms on which such alternative life forms might be based can in no way be used as an argument against their existence (particularly in view of the fact that we do not yet understand all the essential mechanisms in our own case). However, even if we suppose for the sake of argument that such alternative life forms are actually more common than our own, their existence is nevertheless quite irrelevant as far as the preceding line of argument is concerned, because the known fact that we ourselves belong to a star–planet based life system must in any case be taken into account for example by including it (in S) as one of the *a priori* restrictions that should be allowed for in estimating the effective *a priori* probability of obtaining the observed value of the Hubble time. Indeed if the basic theory of stellar structure had been as well understood then as it is today, it would have been possible to use the foregoing considerations to predict the order of magnitude of the cosmological expansion rate that we (as a star–planet based life form) could

expect *in advance* of its observational discovery by Hubble (in much the same way as Gamow actually did predict in advance the order of magnitude of the cosmic background radiation). The fact that we observe a value predictable in advance on the basis of the orthodox theory (B) can hardly be used as a justification for adopting an alternative theory (A) as Dirac wished to do.

Another way of reaching the same conclusion is by operating at an earlier level in the succession of inferences. There is no absolute distinction between the facts that are considered as *a priori* restrictions (S) and those that are considered as *a posteriori* evidence (E): all that matters is that they should be taken into account whenever available and relevant. We could perfectly well choose to operate at a more fundamental level, treating the fact that we belong to a star–planet based life system as *a posteriori* evidence (in E) rather than as an *a priori* restriction (in S). If, as devil's advocates, we were also to adopt the assumption that most intelligent observers in the Universe belong to other quite differently based life forms, the observed Hubble time and the concomitant fact that we belong to a star–planet based system would not be predictable in advance: in this representation, unlike the previous one, we would have $p_S(E/B) \ll 1$. However (unless it were claimed that varying gravitational coupling in some way increases the odds in favour of star–planet based life forms) the defenders of Dirac's theory would also have to face the fact that our own case was exceptional, i.e. $p_S(E/A) \ll 1$, again in contrast with the previous representation. A really persistent devil's advocate might still try to wriggle out of this by suggesting that the odds in favour of star–planet based life forms might be higher within the framework of theory A, but (in the absence of explicit ideas about the nature of the alternative life forms) this objection would carry no weight whatsoever, because the prosecution could make the analogous claim in favour of B with equal plausibility. Thus the final comparison factor, $p_S(E/A)/p_S(E/B) \approx 1$, would still be the same as before.

Although (as we have just seen) it is not directly pertinent to the examples that have been considered above, the conceivable (if not explicitly imaginable) existence of radically different life forms is certainly relevant to applications of what I have called the 'strong anthropic principle', which is the analogue, at a more fundamental level of inference, of the ordinary 'weak' anthropic principle that has been the subject of discussion so far. As I originally formulated it (Carter 1974) this 'strong' principle consisted in the remark that our mere existence as intelligent observers imposes restrictions not just on our situation but even on the general properties of the Universe, including the values of the fundamental parameters that are the subject of the present meeting. Although this 'principle' has aroused considerable enthusiasm in certain quarters, it is not something that I would be prepared to defend with the same degree of conviction as is deserved by its 'weak' analogue.

The strong principle is less satisfactory than the weak one for two distinct kinds of reason. First, it is not so evident that it is really applicable even in the conservative form stated above, because on the one hand it is not clear that the unified theories towards which we are progressing (as described in other contributions to this discussion meeting) will ultimately leave over any parameters that are 'fundamental' in the sense of being independently variable in such a manner as to be meaningfully selectable, and even if so, it is, on the other hand, not clear either whether (in view of our ignorance of alternative life forms) these values really would be restricted in the way one might expect naïvely from our own case, on the basis of examples suggested, for example, by Hoyle (1954), Carter (1974, 1976), Carr & Rees (1979), Barrow & Silk (1980) and Nanopoulos (1980). (General discussions have been given by Davies (1980), Demaret & Barbier (1981), Breuer (1982) and Barrow & Tipler (1983); see also the contributions by Barrow and by

Press to the present discussion meeting.) Second, even if it is valid in the above sense, this 'strong' principle cannot be used to make actual predictions except by selecting a 'cognizable' subset within the framework of a hypothetical (and not particularly plausible) ensemble of universes whose existence is commonly assumed to be included implicitly in a less conservative formulation of the 'strong anthropic principle'.

Even the choice of the term 'anthropic' is less judicious in the 'strong' than in the 'weak' case: in retrospect, I regret not having used an expression, such for example as 'the cognition principle', having a more transcendent connotation. The philosopher Gale has recently gone so far as to suggest (Gale 1981) that (in conjunction with the world ensemble hypothesis) this *cognizability* principle might be promoted to the status of a 'reality' principle, but I would like to dissociate myself vigorously from such a proposition, which apparently reflects a widespread misconception among philosophers to the effect that science is concerned with 'reality', whereas actually (unlike philosophy and theology) science is only concerned with 'realism'. The same misconception would appear to be implicit in the doctrine that scientific theories are never verifiable but only falsifiable. This doctrine would of course be justified if one considered that theories should precisely represent universal 'truth', but by such a standard all existing scientific theories are not only falsifiable, but may safely be assumed in advance to be false, which has the implication that (according to the standards imposed by philosophers!) science has so far achieved nothing at all. In practice, however, science is not concerned with underlying truth but more modestly (and, by its own criteria, more successfully) with providing the most simple, coherent and comprehensive possible description of *appearance* (desiderata such as objectivity being biproducts of these requirements). (For example it is for theologians to decide whether or not the devil created the fossil record as a perfect fake 5986 years ago: the answer to this classic question in no way affects the purely scientific problem of understanding what it was that is represented by the – genuine or fake – record.) Scientific theories should not be judged as true or false, but rather should be evaluated as relatively good or bad on the basis of criteria such degree of accuracy, range of applicability, etc. The best theories can predict results in advance, but even partial historical explanations or mere botanical classification of previously known results should not be dismissed as valueless. Applications of the 'strong anthropic principle' should be judged by the standards of this humbler, merely explicative rather than predictive category.

In the remainder of this discussion we shall only be concerned with applications of the ordinary 'weak' anthropic principle whose genuinely predictive power should become apparent, even if it is not already evident from the examples of the preceding section.

3. The remarkable coincidence between the timescale of past biological evolution on Earth and the future life expectancy of the Sun

I now come to the first significant new point that I wish to make in the present discussion, which concerns the relevance of the anthropic principle for the interpretation of the observational evidence pertaining to biological evolution on Earth by the Darwinian selection process. To start with I would like to draw the attention of biologists to an application (Carter 1982) which is very simple, but not quite so obvious as Crick's 'statistical fallacy' (that was described in §1). This application is based on a hitherto neglected numerical coincidence that I personally

consider to be much more significant than, for example, the large number coincidence of Dirac (that was also described in §1).

The coincidence to which I am referring is based on the very well known fact (see, for example, Dickerson & Geiss 1976) that the time t_e, say, that has been taken so far by biological evolution on this planet since its formation is given to within a few tens of percent by

$$t_e \approx 0.4 \times 10^{10} \text{ years} \tag{3.1}$$

and the almost equally well known fact (see, for example, Hoyle 1955) that the 'main sequence' lifetime, τ_0 say, of the Sun, during which the energy output from steady hydrogen burning can maintain favourable conditions for life on Earth, is estimated to be given with not quite comparable precision by

$$\tau_0 \approx 10^{10} \text{ years.} \tag{3.2}$$

Now the biological processes that have governed the evolution of life up to the present stage of emergence of civilization and the astrophysical processes determining the lifetime of the Sun have nothing directly to do with each other (the slowness of the former arising from the numerical complexity of living systems, whereas the slowness of the latter arises from the weakness of gravitation). Therefore the coincidence of these numbers to within a factor close to two, representing the observation that the Sun is now just about half way through its expected life, does not deserve to be just taken for granted as it seems to have been until now. (Indeed, simply in terms of precision, this coincidence is much more striking than the order of magnitude cosmological coincidences which not unjustifiably caught the attention of Dirac.)

Whereas the principal physical processes governing the lifetime of the Sun are generally believed to be adequately understood, the very complicated mechanisms governing the evolution of living systems cannot yet be analysed, still less predicted, in other than very vague qualitative terms. We certainly do not know enough to predict from first principles whether the expected average time $\bar{t}$ which would be intrinsically most likely for the evolution of a system of 'intelligent observers', in the form of a scientific civilization such as our own, should take much less or much more time than is allowed by the external restraints that limit the duration of favourable conditions. In such a state of ignorance, both of these two alternative possibilities should therefore be retained for consideration as not implausible *a priori*. Only the intermediate borderline case, in which the intrinsically most likely evolution time came out to be of just the same order as the time allowed by external restraints, could be set aside in advance, as being much less plausible *a priori*, and therefore worth considering seriously only if convincing *a posteriori* evidence were obtained against both the other two possibilities.

Now the first of these two possibilities, namely that the intrinsically expected time $\bar{t}$ is very short compared with the externally allowed time τ_0 is indeed rather convincingly excluded by the *a posteriori* evidence that the observed evolution time t_e is not small compared with τ_0, since it is hard to think of any particular reason why our arrival should have been greatly delayed relative to the intrinsically expected time $\bar{t}$. However, provided one avoids making the habitual mistake of overlooking the anthropic principle, it can easily be seen that the observation that t_e is comparable with the upper limit τ_0 is just what would be expected if we adopt the alternative hypothesis that the intrinsically expected time $\bar{t}$ is much longer than τ_0: in this case self-selection ensures that ours must be one of the exceptional cases in which evolution has proceeded much faster than usual; on this basis it is to be expected that t_e should be comparable with τ_0 because there is no particular

[143]

reason why we should belong to the even more exceptional cases in which evolution proceeds even more rapidly although, with the assumption that the Universe is infinite, such cases must of course exist.

Since this satisfactorily accounts for the observed order of magnitude of t_e, there is no need to have *ad hoc* recourse to the *a priori* less plausible hypothesis that the magnitudes of such unrelated quantities as l and τ_0 should just happen to coincide. Although the preceding inductive argument is essentially probabilistic, and hence cannot be absolutely watertight, I consider that it constitutes rather strong evidence for the conclusion that l is in fact much larger than τ_0. In particular, this means that there is no justification for the implicit assumption by many science fiction writers that the expectation l should be comparable with the observed value t_e. Our present conclusion will no doubt be unpopular in such quarters (which perhaps explains why it has not been pointed out before) because it implies as an obvious corollary – which will be discussed in a rather more quantitative manner in the following sections – that civilizations comparable with our own are likely to be exceedingly rare (even if locations as favourable as our own are of common occurrence in the galaxy, which is by no means evident) so that not much credibility can be attached to the exciting fiction scenarios involving reception of extraterrestrial communications, not to mention visitations.

4. Régimes of chance and necessity in biological evolution

Let us now consider the mechanism of biological evolution in rather more detail. According to the memorable description of Monod (1970) this mechanism is governed by a combination of 'chance and necessity'. The 'chance' here refers to the essentially random mutations in the genetic information passed on by an individual to its immediate descendants, while Monod's 'necessity' refers to the ineluctability with which the Darwinian natural selection process can ultimately impose certain particular kinds of mutation on a species as a whole in suitable circumstances. For many practical purposes, however, Monod's 'necessity' is rather illusory: to start with, there is the limitation (which is particularly relevant to laboratory experiments) that it only applies to very large interbreeding populations; of even greater importance for our present purposes is the consideration that even those changes that really are imposed by natural selection will themselves be functions (and often very sensitive functions) of external ecological conditions. The conditions determining the direction of Darwinian selection, even when it is effective, vary in such a complicated manner owing to the interplay with other species, not to mention geophysical effects, that for the practical purpose of evaluating very long term evolutionary trends they must be treated as being to a very large extent governed by chance. In other words, the relevance of Monod's 'necessity' is effectively confined to an intermediate level, between the stochastically governed régime of microscopic mutation processes and the hypermacroscopic régime of very long term evolutionary trends which is also effectively dominated by random effects.

In order to make these ideas rather more explicit, let us consider some very simple mathematical models that illustrate some of the general qualitative features of evolutionary processes, as they have been understood since the work of pioneers such as Wright (1931). To start with, let us consider a hypothetical biological species in which the hereditary genetic information content carried by each individual is I bits, and in which each separate binary unit is subject to independent random mutations whose rate (per individual per generation) is given by μ/I where the

quantity μ so defined is what we shall refer to as the mutation rate factor, and typically can be expected to be very roughly of the order of unity, i.e.

$$\mu \approx 1. \tag{4.1}$$

(It is generally accepted that the genetic information content of existing terrestrial species is carried by nucleic acid chains with a language in which the basic letters are twenty kinds of amino acid that are specified (with some redundancy) by a 6-bit code (in much the same way as is done for the symbols of an ordinary typewriter by the 8-bit ASCII code in modern computer information processing.) If one needed a more precisely realistic model, one would have to allow for the fact that information is duplicated so as to allow sexual interchange, and also that the mutations are not completely independent – the most elementary mutation involving not 1 but 2 bits, corresponding to the four different kinds of nucleotide. However, these complications are unimportant for the crude estimates that will be made here.)

The reason why the rate factors μ can be expected typically to be of the order of unity is that their average value $\bar{\mu}$, as taken over all the I distinct rate factors, is just the *total* rate (per individual per generation) of binary mutations of all kinds occurring in the model. Now this total mutation rate $\bar{\mu}$ cannot be very large compared with unity if the species is to survive because a significant proportion of all mutations can be expected to be lethal. On the other hand it would be disadvantageous for long term adaptability if $\bar{\mu}$ were unnecessarily low. One is therefore led to expect that the internal mechanisms controlling the mutation rate would adjust themselves so as to give values of $\bar{\mu}$ at least roughly of the order of unity, a prediction which has been investigated by a number of workers, and which seems to have been confirmed in many cases. (This reasoning also entails that there should only be rather limited scope for acceleration of evolution by an externally induced increase in mutation rates such as might be induced by exposure to intense radiation.) The total relevant genetic information content I is hard to evaluate experimentally: measurements of total nucleic acid content give reasonably precise upper limits, ranging from the order of 10^9 for bacteria to the order of 10^{10} for mammals, including ourselves (see, for example, Dobzhansky *et al.* 1977), but there is reason to believe that a considerable fraction of this is effectively redundant, so that somewhat lower estimates of I would probably be more appropriate.

Under sufficiently favourable conditions, which include the requirement that the effective interbreeding population number, N say, be large enough, the changes that are favoured by natural selection will be imposed not just 'necessarily' (in Monod's sense) but indeed very rapidly compared with geological timescales. It is easy to see that if a particular kind of mutant is relatively favoured in its breeding by a selection coefficient s (meaning that if no new mutations occurred the mutant fraction, q say, would increase by a factor e^s from one generation to the next) then even if it started from nothing the mutant fraction q would increase so as to reach the order of unity within a timescale T whose order of magnitude, in units of a generation, is given by

$$T \approx s^{-1} \ln (s\mu^{-1}I), \tag{4.2}$$

provided the two following conditions are satisfied. First we must have

$$s \gtrsim \mu I^{-1} \tag{4.3}$$

in order for Darwinian selection to be able to dominate the average mutation rate, and second we require

$$N \gtrsim \mu^{-1}I, \tag{4.4}$$

[145]

which is sufficient to ensure that random breeding fluctuations can be neglected. Subject to a more severe restriction on the breeding selection coefficient, s, namely

$$N\mu I^{-1} \gtrsim s \gtrsim N^{-1}, \tag{4.5}$$

the expression (4.2) will remain applicable even for comparatively small populations in the range

$$N^2 \gtrsim \mu^{-1}I \gtrsim N. \tag{4.6}$$

If it is assumed that s is not too small and μ not too large compared with unity, their contributions inside the logarithm will be unimportant, and thus the expression (4.2) for what may be thought of as the typical number of generations required for changes by ordinary Darwinian selection effectively reduces simply to $T \approx s^{-1}\ln I$. In view of the fact that sexual interchange within the population makes it possible for many independent kinds of mutation to be undergoing selection in this way all at the same time, one sees that it is in principle possible to make a transition between any two among the 2^I possible information configurations (of which most are of course not viable) within a timescale which depends merely logarithmically on I. Since $\ln I$ is at most in the region of 20 for the human population, we see that even in our own case comparatively modest selection factors, of the order of 1 % or so, could bring about major evolutionary changes within a few thousand generations, i.e., in substantially less than 10^5 years. (On the supposition that $\mu^{-1}I$ is in the vicinity of 10^9, the prehistoric human population would probably have been within the range (4.6) most of the time, while the present population has just about reached the range (4.4).)

For extremely small populations the genetic evolution is no longer subject to a régime of 'necessity' but to a régime of 'chance'. Even for quite high values of the Darwinian selection coefficient, s, the initiation of the quasi-deterministic selection process characterized by formula (4.2) will be subject to a stochastic delay, with a characteristic mean timescale τ given in order of magnitude by

$$\tau \approx N^{-1}\mu^{-1}I, \tag{4.7}$$

which results from the fact that selective breeding cannot begin to operate until the first appearance of mutants of the kind under consideration. Once a few of these mutations have occurred, after an average time given by (4.7), selective breeding will establish the mutant strain comparatively rapidly provided that the condition

$$s\mu^{-1}I \gtrsim N \gtrsim s^{-1} \tag{4.8}$$

is satisfied. However, if the selective breeding coefficient, s, is too small, another element of 'chance' enters the situation, in so much as random fluctuations (proportional to $N^{\frac{1}{2}}$) in relative numbers of descendants will dominate: thus subject to the necessary conditions

$$\mu^{-1}I \gtrsim N, \quad s^{-1} \gtrsim N, \tag{4.9}$$

one obtains a 'neutral' régime of what may be described as stochastic genetic drift in which (as described, for example, by Kimura 1979) mutant strains are imposed on the population, or eliminated, at random with a characteristic timescale τ for any particular such kind of mutation given by

$$\tau \approx \mu^{-1}I. \tag{4.10}$$

(This is the mechanism that is thought to be responsible for the minor differences between analogous proteins in related species, which have been used as a genetic clock for establishing the

comparative lengths of the branches of phylogenetic trees.) A similar timescale is obtained for 'neutral' genetic drift in large populations when the selective breeding coefficient s is negligible, which arises when

$$N \gtrsim \mu^{-1}I, \quad s^{-1} \gtrsim \mu^{-1}I; \tag{4.11}$$

but in this case the process is not effectively stochastic, in so much as the mutant fraction q will 'necessarily' increase, slowly and steadily, until it reaches a value of order unity after a time T given by

$$T \approx \mu^{-1}I, \tag{4.12}$$

as a result of simple linear accumulation of mutations in succeeding generations.
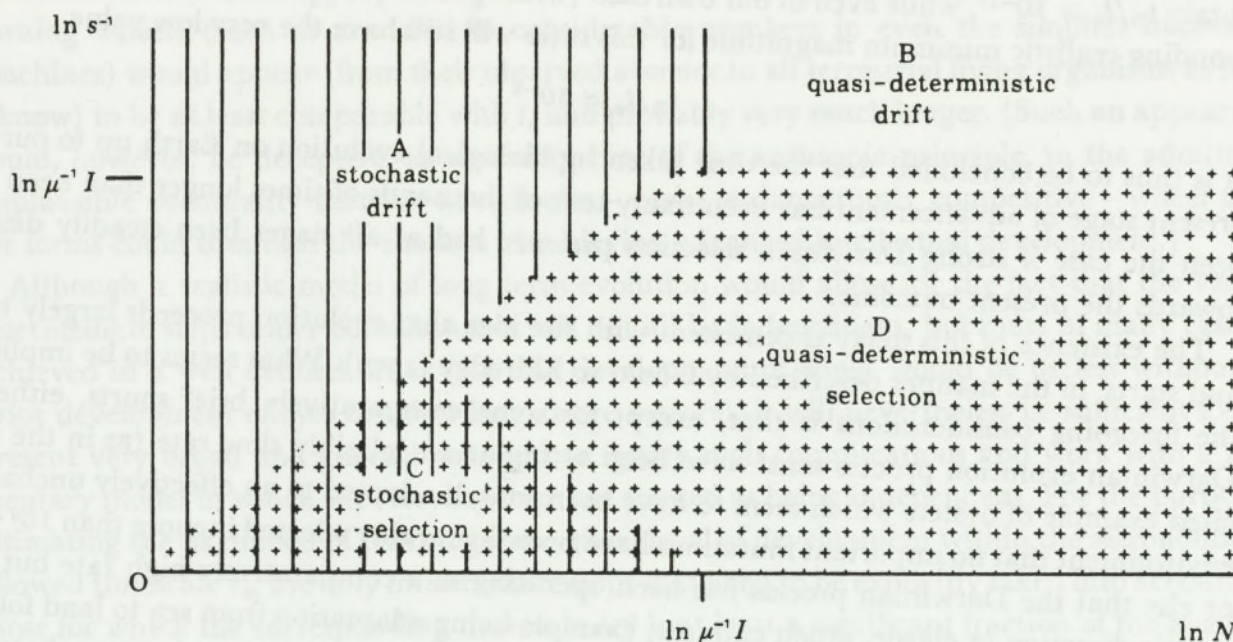

FIGURE 1. The approximate locations of régimes of 'chance' and 'necessity' are shown on a plot of the inverse of the selective breeding coefficient s against the effective breeding population number N with a logarithmic scale normalized with respect to the critical magnitude $\mu^{-1}I$ which represents the inverse of the relevant mutation rate. Darwinian régimes in which natural selection is dominant are marked by crosses, $+ + +$, and stochastic régimes in which the population is too small for Monod's 'necessity' to apply are marked by vertical shading |||. The four possible combinations are: A the régime of neutral random drift (whose significance was pointed out by Kimura) with characteristic time given by (4.10); B the régime of very slow quasi-deterministic drift on a timescale of the same order, as given by (4.12), C the régime of Darwinian selection in a small population at a stochastically variable rate with characteristic time given by (4.7); finally D the régime of ordinary quasi-deterministic Darwinian selection in a large population, on a potentially short timescale given by (4.2).

The relative positions of the four evolutionary régimes that have just been outlined are plotted on a $\ln N / \ln s$ plane in figure 1.

5. A SIMPLE STOCHASTIC MODEL FOR THE ERRATIC PROCESS OF LONG TERM EVOLUTION

The timescale given by both (4.10) and (4.12) for unselected 'neutral' evolution is very long and indeed in our own mammalian case is comparable with the observed total biological evolution time t_e (which is itself a rather striking coincidence for which I can at present think of no plausible explanation, anthropic or otherwise, in view of the fact that such undirected

evolution can hardly have been the primary mechanism in our own emergence even though it must presumably have had a certain subsidiary role). The timescale (4.7) for the stochastic régime of Darwinian selection in small populations will usually be much shorter, but it too is limited by the necessarily low rate, μI^{-1}, of mutations of a given kind per individual per generation.

In contrast with these, the timescale given by (4.2) for ordinary quasi-deterministic Darwinian selection in a large population is virtually independent of the mutation rate. In ordinary time units, as opposed to generations, the corresponding time, t_D say, is expressible very roughly as

$$t_D \approx \tau_g s^{-1} \ln I, \tag{5.1}$$

where τ_g is the average time interval between generations. This time τ_g can be extremely short compared with the observed total evolution time t_e, even for quite moderate values of the selective breeding coefficient s. Thus for bacteria, (with τ_g of the order of a few hours) one can easily obtain $t_D/t_e \approx 10^{-12}$, while even in our own case (with τ_g of the order of tens of years) the corresponding realistic minimum magnitude for this ratio will still have the very low value

$$t_D/t_e \approx 10^{-5}. \tag{5.2}$$

It is thus to be concluded that the time taken by biological evolution on Earth up to our own present stage of advancement has been many tens of thousands of times longer than need have been the case if strong Darwinian selection pressure had at all stages been steadily directed towards the present outcome.

The existence of this delay is consistent with the idea that evolution proceeds largely by fits and starts, in the manner described by Gould & Eldredge (1977). What seems to be implied by the foregoing considerations is that, except for some comparatively brief spurts, either the Darwinian evolution process must have been acting at a negligibly slow rate (as in the much cited example of oysters which seem to have been so well adapted to an effectively unchanging environment that no important mutations have been favourably selected in more than 10^8 years) or else that the Darwinian process has been operating at a comparatively high rate but in an erratic direction (a classic, albeit extreme, example being adaptation from sea to land followed by readaptation to sea). This is to be expected from the consideration that the strongly selected changes (describable in terms of Monod's necessity) that must have occurred at many intermediate stages were not teleologically directed towards our present state or any other long term goal but were directed towards immediate advantages in stochastically changing environmental conditions (the changes in question being largely due to complicated interactions with other species, as well as geophysical effects). Thus, to sum up, one can envisage evolution in terms of comparatively rapid adaptation to niches which are themselves undergoing more or less rapid variation in the space of ecological parameters.

This reasonably well established conclusion brings me to the point at which it is appropriate to introduce further simplifying hypotheses to enable us to set up another rudimentary mathematical model for the purposes of describing evolution up to some particular given state of 'advancement', which we shall later on take to be the formation of a scientific civilization such as our own. This model is based on the supposition that the attainment of a given degree of development of a particular kind (e.g. intellectual, in the application we have in mind) depends on successful passage through a number of intermediate steps involving the acquisition of relevant accessories (e.g. eyes). We further suppose that even in globally favourable circumstances the effective unpredictability of local ecological conditions is such that passage through such a step

does not take place automatically but occurs only with a certain probability, λ say, per unit time. The rate λ at which such an intermediate step is likely to occur will depend on the sensitivity of its selective value to variations in ecological conditions. For example, the selective value of even very rudimentary optical radiation detectors would appear to be sufficiently universal to ensure that the appropriate timescale λ^{-1} for the development of eyes is quite short compared with the observed age t_e of the Earth. On the other hand, the acquisition of wings (which in any case is apparently unnecessary for the attainment of our present stage of advancement) would seem to be rather more difficult, having been achieved in only a small number of independent lines of descent, which suggests that such a development should be described in terms of a substantially smaller value of λ (so that λ^{-1} would be of almost comparable magnitude with t_e), due presumably to the fact that rudimentary flying equipment is usually a burden rather than an advantage. As a final example, the appropriate value of λ^{-1} for the biological development of ordinary freely turning wheels (such as are used in considerable numbers in even the simplest man-made machines) would appear (from their observed absence in all terrestrial living organisms as far as I know) to be at least comparable with t_e and probably very much longer. (Such an appearance could, however, be deceptive, as a consequence of the anthropic principle, in the admittedly implausible eventuality that the advent of a – perhaps dangerously competitive – wheel using life forms could diminish the chances of achieving advanced intellectual development.)

Although a realistic model of long term evolution would allow for the fact that the various intermediate steps under consideration will not all be independent, but must in many cases be achieved in a well defined order (e.g. the development of wings would be useless without the prior development of eyes or other long range sensors), it will nevertheless be sufficient for our present very broad and general concerns to ignore such complications and work with a rudimentary model in which the relevant steps are treated as being independent. For the purpose of estimating the likelihood of reaching some given level of development within the astronomically allowed timescale τ_0, the only intermediate steps that need to be explicitly taken into account are those for which the corresponding timescale λ^{-1} is at least a significant fraction of the available time τ_0, since it can be taken for granted that all the others will be achieved with virtual certainty. Thus there will remain only a limited number, n say, of 'critical' steps that are characterized roughly by

$$\lambda^{-1} \gtrsim \tau_0, \tag{5.3}$$

and which must therefore be taken into account. (It follows from the considerations of the preceding section that n must be at least equal to 1.) For the crude qualitative conclusions that we wish to draw here, it will be sufficient, as a final simplification, to treat the n critical steps that are retained for consideration as each having the same 'average' value of λ. In this simple model, there will be an intrinsic probability given by

$$p = (1-e^{-\lambda t})^n, \tag{5.4}$$

for all n critical steps to have been carried through prior to any given time t. (The application of this model is justifiable provided t is large compared with the characteristic times λ^{-1} of all the non-critical steps that were left out of account, and hence in particular whenever t is at least comparable with τ_0).

Having thus set up the simplest plausible model for the description of long term evolutionary

'progress', let us now consider its implications. To start with we can evaluate the average time $\bar{t}$ by which all the n critical intermediate steps are completed: substitution of (5.4) gives

$$\bar{t} = \int_0^\infty t \, \mathrm{d}p = \lambda^{-1} h_n, \tag{5.5}$$

where

$$h_n = 1 + 1/2 + \ldots + 1/n \tag{5.6}$$

is the Euler sequence, which for large values of n is given asymptotically by the expression

$$h_n = \ln n + \gamma + O(1/n), \tag{5.7}$$

where $\gamma = 0.577\ldots$ is Euler's number.

In applying the model to the emergence of a scientific civilization on Earth, we recall the considerations of §3 which suggest that we must have

$$\bar{t} \gg \tau_0, \tag{5.8}$$

since to have $\bar{t} \approx \tau_0$ would *a priori* be an unlikely coincidence; the remaining alternative, $\bar{t} \ll \tau_0$ is hardly compatible with the observation that the actual time t_e of our emergence satisfies the relation

$$t_e \approx \tau_0. \tag{5.9}$$

Although no obvious external considerations can affect the intrinsically low probability,

$$p \approx \exp\{-h_n t_e / \bar{t}\} \tag{5.10}$$

of obtaining $t_e \gg \bar{t}$, on the other hand the upper limit τ_0 on the timescale during which the conditions for applicability of the model are maintained makes it certain, by application of the (ordinary, weak) anthropic principle that we shall have $t_e \lesssim \tau_0$ even if (5.8) is satisfied.

In view of the fact that the distribution $\mathrm{d}p/\mathrm{d}t$ obtained from (5.4) increases monotonically up to a maximum at a time t_m given by

$$t_m = \lambda^{-1} \ln n \approx \bar{t}, \tag{5.11}$$

we see that subject to the hypothesis (5.8) the cut-off, $t_e \lesssim \tau_0$, leads to the prediction that the relative probability of obtaining the observed coincidence (5.9) is of order unity. As an immediate corollary of the same observationally consistent hypothesis (5.9) we obtain the prediction that the probability of completing all the n necessary intermediate steps on a randomly chosen planet subject to the same average global environmental conditions as our own will have the very small value

$$p \approx (h_n \tau_0 / \bar{t})^n, \tag{5.12}$$

since for any reasonable value of n, the order of magnitude of the factor h_n cannot be very large compared with unity. Quite moderate lower bounds $n \gtrsim 10$ and $\bar{t}/\tau_0 \gtrsim 10^2$ give $p \lesssim 10^{-20}$, which is more than sufficient to ensure that our stage of advancement is unique in the visible Universe. In view of our utter lack of quantitative understanding of the mechanisms that determine the rates λ, it is by no means implausible that the relevant mean time should have a very much large value, for example $\bar{t}/\tau_0 \gtrsim 10^{10}$, which would be sufficient to ensure that our level of advancement is unique in at least our own galaxy even if the number n of necessary critical steps is as low as one.

6. THE NUMBER OF CRITICAL STEPS: A DILEMMA?

The preceding interpretation of the conclusions of §3 within the framework of our simplified stochastic model does not exhaust the conclusions that can be drawn from the present application of the anthropic principle. We shall now go on to point out some further implications of a less trivial and more interesting nature.

The power law dependence,

$$p \approx \alpha t^n \tag{6.1}$$

(for constant α) which is obtained from (5.4) in the limit for $t \ll \bar{t}$ (and which would *still* hold for a more sophisticated model taking account of differing rates λ from one step to another and of the necessity that the critical steps be taken in the correct order) implies that with a relative probability close to unity the completion of the n critical steps within the allowed time range $0 < t < \tau_0$ will occur near the end of this range to within a fraction of the order of magnitude of $1/n$. Thus due allowance for the anthropic principle not only explains the order of magnitude relation (5.9) but also leads to the much more precise prediction

$$\tau_0 - t_e \approx n^{-1}\tau_0. \tag{6.2}$$

Now, as we have already remarked, the observational coincidence (5.9) is valid not merely as a crude order of magnitude relation but within a factor close to two, in so much as the standard estimates of the duration of the future main sequence life of the Sun give a value of the same order as the age of the Earth, i.e.

$$\tau_0 - t_e \approx t_e, \tag{6.3}$$

to within an uncertainty factor hardly exceeding 100 %. Thus provided the broad theoretical framework we have been using is essentially correct (and it is to be emphasized that the reasoning in this section remains valid even when one allows for ordering and unequal probabilities of the critical steps) one sees that in order to fit the prediction (6.2) with the observation (6.3) we must have at most a very small number of critical steps: the values $n = 1$ and $n = 2$ are quite consistent, but values from $n = 3$ onwards become rapidly more difficult to reconcile with the comparatively long period during which terrestrial conditions seem likely to remain favourable. In short, combining this conclusion with that of the previous section, we are led to induce that

$$1 \leqslant n \lesssim 2. \tag{6.4}$$

My first reaction on arriving at this quite severe upper limit on n was one of surprise, since I had previously been inclined to think that the appropriate value of n (i.e. the number of intermediate steps necessary for evolution to our present stage whose rates λ are low on timescales comparable with the age of the Earth) was likely to be very large. This vague prejudice arose from a certain appearance in the fossil record of a consistent trend in the direction of long term evolution towards our present state (the same appearance that in an extrascientific context has often been used as an argument in favour of the idea of 'divine guidance'). One has the impression that successively more recently evolved categories of animals can be classified in a rising hierarchy of levels of increasing 'advancement', this latter term being defined in terms of acquisition of features apparently necessary for the ultimate emergence of civilization. The transitions between these levels would be candidates for treatment as critical steps in the sense of our simple stochastic model, a typical example being the development of the placenta, for which the corresponding rate λ must be fairly low in view of the fact that it has occurred only on one side of the Wallace line. The more or less steady occurrence of progressive steps such as this every 10^8 years

[151]

or so could be accounted for in a satisfactory manner in terms of our simple model by taking a value of n of the order of at least several tens. However, the appearance of relatively steady progress up the evolutionary ladder is harder to account for in terms of a model in which the value of n is very small.

I can at present imagine two quite distinct ways of resolving the dilemma that has just been posed. The first is to suppose that the appropriate value of n really is large (as is suggested by the immediately preceding considerations) and that the prediction (6.2) (on which the upper limit in (6.4) was based) is invalid due to an overestimation of the available timescale during which geophysical conditions will remain favourable. The example of the phenomenon of ice ages shows that the terrestrial climate is sensitive to factors that are still not well understood, so that it is difficult to exclude the possibility that our environment was destined (had we not emerged to retard or accelerate the process) to become unfavourable due to some as yet unforeseen over-heating or overcooling effect within a timescale, τ_e say, only very slightly greater than the present age, t_e, of the Earth. On this hypothesis, our previous line of reasoning would need to be modified by the substitution of τ_e in place of τ_0 in (6.2).

Instead of resorting to such an *ad hoc* hypothesis, the alternative way to resolve the dilemma would be to adopt the conclusion (6.3) bravely at its face value, and to submit the rather tenuous reasons for doubting it to closer scrutiny. This means accepting that at most one or two of the steps in our evolution (e.g. the original establishment of the genetic code, and the final break-through in cerebral development) were genuinely critical in the sense of our stochastic model. The implication is that all the other apparently important and not obviously ineluctable steps (such as the development of the placenta), are either less difficult than one might suppose or else are merely incidental and not as essential as is widely believed. My present inclination is to believe that the latter is the most likely. In other words, my best guess is that the correct conclusion to be drawn from the reasoning that has been presented is that many of the salient developments in our evolution were quite unnecessary, as well as having been intrinsically improbable, in so much as many alternative evolutionary pathways would have been compatible with the ulti-mate emergence of civilization (e.g. there is no obvious reason why it should not have arisen in egg laying animals).

If this last interpretation is correct, it means that the apparent existence of an evolutionary ladder is to a large extent an illusion: an artefact of our still unduly anthropocentric imaginations, which lead us to jump too easily to the conclusion that merely because we happen to possess some particular attribute it must be essential for 'higher development'. (If we had happened to be born – or for that matter, hatched – with wings, they would no doubt be generally regarded as an indispensable status symbol for any life form aspiring to be described as 'advanced'!) This final remark leads me to recapitulate the general moral of this exposition, which is that one should try to steer a moderate course between the Scylla of excessive anthropocentrism and the Charybdis of unjustifiable neglect of anthropic selection effects.

REFERENCES

Barrow, F. J. & Tipler, F. 1983 (In preparation.)
Barrow, J. D. & Silk, J. 1980 *Scient. Am.* **242**, 127.
Bondi, H. & Gold, T. 1948 *Mon. Not. R. astr. Soc.* **108**, 252.
Breuer, R. 1982 *Das Anthropische Prinzip.* Wien: Meyster.
Carr, B. J. & Rees, M. J. 1979 *Nature, Lond.* **278**, 605.

Carter, B. 1974 In *I.A.U. Symposium* **63**: confrontation of cosmological theories with observational data (ed. M. S., Longair), p. 291. Dordrecht: Reidel.

Carter, B. 1976 In *Atomic physics and fundamental constants* **5** (ed. J. H. Sanders & A. H. Wapstra), p. 650. New York: Plenum Press.

Carter, B. 1982 *Pour la Science* **52**, 55.

Crick, F. 1981 In *Life itself: its origins and nature*. p. 90. New York: Simon and Schuster.

Davies, P. 1980 *Other worlds*. London: J. M. Dent.

Demaret, J. & Barbier, C. 1981 *Rev. Quest. Sci.* **152**, 461.

Dicke, R. H. 1961 *Nature, Lond.* **192**, 440.

Dickerson, R. E. & Geiss, I. 1976 *Chemistry, matter and the Universe*. Reading, Massachusetts: Benjamin.

Dirac, P. A. M. 1937 *Nature, Lond.* **139**, 323.

Dobzhansky, T., Ayala, F. J., Stebbins, G. L. & Valentine, J. W. 1977 *Evolution*. San Francisco: Freeman.

Gale, G. 1981 *Scient. Am.* **245**, no. 6, 114.

Gould, S. J. & Eldredge, N. 1977 *Paleobiology* **3**, 115.

Hawking, S. W. 1982 *Physics Lett.* B **115**, 295.

Hoyle, F. 1949 *Mon. Not. R. astr. Soc.* **109**, 365.

Hoyle, F. 1954 *Appl. J. suppl.* **1**, 121.

Hoyle, F. 1955 *Frontiers of astronomy*. London: Heinemann.

Kimura, M. 1979 *Proc. natn. Acad. Sci. U.S.A.* **76**, 3440.

Monod, J. 1970 *Le hazard et la necessité*. Paris: Seuil.

Nanopoulos, D. V. 1980 *Physics Lett.* B **91**, 67.

Press, W. H. & Lightman, A. P. 1983 *Phil. Trans. R. Soc. Lond.* A **310**, 323.

Salpeter, E. E. 1957 *Phys. Rev.* **107**, 506.

Wright, S. 1931 *Genetics, Princeton* **16**, 97.

Discussion

W. H. McCrea, F.R.S. (*University of Sussex, Brighton BN1 9QH, U.K.*). In Dr Carter's scheme for the evolution of man on Earth, it appears that some form of catastrophic happening plays an essential role. What would the outcome be were such a happening not forthcoming? I ask because it seemed almost as though evolution had to anticipate a catastrophe.

B. Carter. I prefer to use the more neutral term 'cut-off' rather than 'catastrophe' for the natural astrophysical (or geophysical) time limitations in question. (Even effects like the ice age phenomenon occur on timescales so long compared with those characteristic of technical develop-ment of civilization that there would be adequate time to prepare countermeasures of avoidance or protection. The term catastrophe would be more appropriate for something like a man made ecological disaster, which is an eventuality that might well be discussed with reference to the anthropic principle, but which has nothing directly to do with the considerations that I have presented on this occasion.) Although the certainty of a long term astrophysical cut-off (and the conceivable possibility of a geophysical cut-off in a less distant future) played an essential role in the foregoing line of argument, such a (future) cut-off evidently was not essential for our (past) development as such. What would occur in an imaginary universe that started off in the same way as our own, but in which such cut-offs were miraculously suspended, is that advanced life would ultimately become far more common than it seems to be in our own Universe now, at least in so far as its density was measured in civilizations per galaxy. However, although there would be a very few rare exceptions that would come into existence under conditions similar to our own, most of the life systems in such a universe would not produce anything describable as a scientific civilization until after the universe had reached such an advanced state of expansion that the average number of civilizations per unit volume of intergalactic space might actually be lower than in our own Universe now (particularly if the number n of critical steps really is less than three as has been suggested).