Some Philosophical Issues in Cosmology

Since cosmology is, quite literally, the attempt to apply physics to everything in sight (and much that is not!), it is not surprising that, more than any other branch of science, it raises precisely those questions – about the nature of "explanation", the sense of past and future, etc. – that we think of as typically "philosophical". In discussing them, one should bear a few caveats in mind. First, as already emphasized, in applying the laws of physics as they have been discovered in the laboratory to the Universe as a whole, we are often extrapolating them by not just orders of magnitude but *decades* of orders of magnitude; this is particularly true of discussions of the "early Universe", where according to the standard scenario, the typical energies involved are far beyond anything we can investigate in laboratory experiments now or in the foreseeable future. This point applies even more to the "subjects" of quantum gravity and quantum cosmology, which at present have no experimental basis at all – a fact that it is well to remember when deriving grandiose conclusions from them! Second, even the leading (theoretical) practitioners tend on occasion to be quite naïve when they leave the secure ground of the mathematics of their theories and start discussing philosophical issues; this particularly applies to discussions of time symmetry in the Universe. Third, the philosophical impact of a new idea in cosmology may be in no way commensurate with its technical importance; an example is the inflationary scenario, which, as a solution of a number of pressing technical problems, is a brilliant innovation, but in the end does not (in my opinion) affect the philosophical issues connected with the Big Bang in any very substantial way.

In this lecture, I will discuss five topics in cosmology that give rise to philosophically interesting questions: the Big Bang (and Big Crunch), "other universes", the anthropic principle, quantum cosmology and the cosmological arrow of time. To some extent, they are all interrelated; so that, for example, one's views on quantum cosmology are likely to be to some degree affected by one's reaction to the anthropic principle as an explanatory device.

The Hot Big Bang

Let's start by reminding ourselves what the Hot Big Bang is. According to the standard FRW (Friedmann-Robertson-Walker) cosmology (and, according to Penrose and Hawking, to generalizations of it that allow for spatial inhomogeneity, etc.), the Universe is described, at given cosmic time t, by a scale factor $R(t)$ (which for a closed Universe is its radius), whose time-dependence is such that if we extrapolate it backward in time, we inevitably reach a singularity – that is, a point at which R tends to zero (and its time derivative becomes infinite). This conclusion appears to hold ineluctably so long as $R(t)$ is described as a classical variable and certain very general conditions hold; it is not affected qualitatively by the occurrence of an inflationary period, if such occurred. (We can still extrapolate the general FRW scenario back to the period before inflation!) However, if we take seriously the idea that everything, including the structure of the Universe itself, should in principle be described in quantum-mechanical terms, then we do eventually get back to a point at which all bets are off because the quantum fluctuations in $R(t)$ become comparable to the quantity itself. This happens when $R(t)$ is "shrunk" down to the so-called Planck length, $\sim 10^{-33}$ cm, and happens at a "time" of the order of 10^{-43} secs relative to the classical singularity. Since no one really knows, at present, how to construct a consistent theory of quantum gravity in a form that would apply under these strongly fluctuating conditions, any attempt to infer the history earlier than this must be based entirely on speculation. For the moment, let me ignore this complication and pretend that we can in fact extrapolate the classical scenario right back to the singularity, and moreover, that there are no physical "events" before this (but see "Other universes", below).

The first obvious (naïve) question is: did the Universe have an origin in time? If so, what went before and what caused it? The first, rather trivial, point that needs to be made is that the conclusion that the classical equation had a singularity somewhere between 7 and 15 billion years ago depends, of course, on our definition of "year", and more generally on our reckoning of so-called "cosmic standard time". If we think about the latter, we immediately have a prima facie problem. Suppose, for example, that we take as our standard of time, as we would tend to do nowadays, some atomic unit such as the period (inverse frequency) of a particular atomic transition, say in cesium. In applying this standard to cosmology, the problem is that until fairly late in the history of the Universe, there actually were no Cs atoms around! (Most heavy elements are believed to form in supernova explosions, and this requires the star in question not only to have been formed, but to have been around for a while.) Even if we use a hydrogen transition as our standard, H atoms were not around, at least in any substantial quantity, until the epoch of "recombination" ($\sim 10^5$ yrs after the HBB). Perhaps we could get around this difficulty by basing our time standard on something more "elementary", e.g., the energy needed to create an electron-positron pair, or the lifetime of the free neutron. But we still have a problem, because as we go back in time toward the HBB, matter becomes more and more densely-packed, and eventually we reach the point where the energy (lifetime) in question depends so strongly on the local environment that it is no longer useful as a standard. Thus, the problem of constructing a physical "clock" that would in principle enable us to *measure* the time development of the Universe in its early stages is nontrivial and indeed perhaps insoluble. Does this matter? Probably one's answer will depend on the extent to which one is prepared to regard the impossibility of constructing such a clock as merely "contingent" – in effect, an annoying but inessential feature of the early Universe – as distinct from something that is quite generic and model-independent.

Even granted that we have found some way of extrapolating our current standard of time back to the earliest days, we are of course perfectly free not to use it. We could simply introduce a new definition of cosmic time, t_{new}, such that t_{new} is some function of t_{old}. A particularly attractive choice might be

$$t_{new} = t_0 ln[(t_{old} - t_{HBB})/t_0]$$

where t_0 is some fixed reference time – let us say, for definiteness, the time at which half the hydrogen in the Universe has recombined – and t_{HBB} is calculated in classical theory. The "advantage" of this choice is that "time" now runs from $-\infty$ to ∞, the HBB occurring at $t = -\infty$! Of course, we pay a price: with such a definition, the laws depend explicitly on time, and in particular, the frequency of a given atomic transition (in the "late" Universe) turns out to be approximately proportional to the time since the HBB. Assuming that our definitions of length are the usual ones, the speed of light would then vary with cosmic time in the same way. With this convention, there is *by definition* no time "before" the Big Bang!

Suppose we adapt the normal convention. Was there then time "before" the Big Bang? This question is somewhat reminiscent of the question: if the Universe is closed, is there "space" outside it? Here we go back to the old Leibniz-Clarke controversy about the notion of space and time. A Leibnizian would presumably say very definitely: Since "space" and "time" are just relationships between events, the whole idea either of space outside a closed Universe, or of time before it "started," is meaningless, since there are no "events" to be related! On the other hand, a Newtonian would presumably say: yes, there was time before the HBB, just as there is space outside the closed Universe. (But this would really only make sense in an "embedding" model; see below.) *If* one takes this point of view, then one can be hit with the Leibnizian (ad hominem) questions: OK, so *why* did the Universe come into being just when it did, and at exactly the place it did? – questions that, even if they are indeed meaningful, current ideas in cosmology certainly provide no answers to.

Does it make sense to ask what "caused" the Big Bang? In particular, if we find that that question has no answer within our current understanding of the laws of physics, does that strengthen the argument for a divine Creator? The problem here is that our normal notion of "cause" implies that when we ask for the "cause" of event B, occurring at time t_B, the answer should be an event A occurring at a time $t_A < t_B$. But whether or not there is "time" before the Big Bang, there are certainly (in the standard picture) no "events" – at least not of any nature describable by physics. In Newton's original thinking, the question presumably does make sense, and the answer might go something like this: Not only was there time before the Big Bang (and space, the "sensorium of God"), but there could exist, at least in principle, "events," namely those of a nonphysical nature associated with the activities of God. Then at some definite time, God took the decision to create a physical Universe; this time just happens to be $\sim$ BCE -10^{10} by our human clocks (but presumably there is some nonphysical "clock" by which it might be something quite different!). The "cause" of the Big Bang was then indeed the action of a divine Creator.

This "Newtonian" view seems to be self-consistent within its limits, but it rests crucially on the notion that there is, as it were, something else beyond the physically observable Universe that has "reality" equal to it. Once one rejects this hypothesis, it is not at all clear that the question "what caused the Big Bang?" can be given any clear meaning, so that the fact that physics cannot answer it does not indicate that we need to invoke something else (any more than the fact that the question "have you stopped beating your wife?" has, in certain circumstances, no answer, by itself implies that tougher laws are needed against domestic violence!).* The question "is there *something else* besides the Universe we know?" leads us naturally to the issue of "many universes", to which I now turn.

Other universes

To discuss the question of whether the concept of "other universes" besides our own is meaningful, we first need to decide exactly what we mean by "our own universe." It is clear that this latter includes all the matter (and radiation, etc.) that we can currently observe, but we usually do not limit the definition to that: crudely speaking, we might try to define "our own Universe" as the sum total of matter (and/ or events) that could conceivably have in any way influenced us (or our ancestors) in the past and/ or may influence us or our descendants in the future. (Clearly, we want to generalize the definition of "ancestor" and "descendant" so that we do not assume that the human race as such is eternal!) Actually, even this definition is not quite adequate (i.e., it does not correspond to the intuitive concept that we are trying to express), since in a closed Friedmann Universe it can be shown that there are some events (on the "far side" of the Universe, as it were) that will remain spacelike-separated from us right up to the Big Crunch, and therefore cannot (according to the postulates of SR) affect us in any way. So we had better make the definition transitive – i.e., postulate that if events A and B are part of the same Universe, and B and C similarly are parts of the same Universe, then by definition, A and C are also parts of the same Universe. It then follows that any event that can be linked with us by a series of "causal" chains, with no restriction on the "direction" of the chains, is part of our Universe even though it may remain forever spacelike-separated from us.

With this definition, let us consider the possible concept of "other" universes besides our own. It is convenient† to classify them into three kinds, according as they are separated from our own in time, space or some other "dimension". To introduce the first kind, let us ask: what are the possible fates of a closed Friedmann universe (such as ours may possibly be) at the final singularity known as the "Big Crunch"? (We could ask a similar question about the Big Bang, but for our purposes it is convenient to do it this way round.) There appear to be three possible answers: (1) "after" the Big Crunch, there is nothing at all; (2) after the Big Crunch, there are further events, but they are causally completely uncorrelated to what happened before; or (3) after the Big Crunch, there are further events, which are causally correlated to what happened earlier. In case (1), the question of other (temporally-separated) universes obviously does not arise. In case (3), according to our definition, it would be "our" Universe that continues through the Big Crunch, and indeed the latter name might be judged inappropriate, since all that would happen is that the Universe would be severely "squeezed" and then released (it is not at all inconceivable that, for instance, a proper treatment of quantum gravity might lead to a result of this nature). In the present context, it is case (2) that

*Cf. Grünbaum's analogy of the Aristotelian and Galilean responses to a question concerning the "cause" of free uniform motion (Leslie, *Physical Cosmology and Philosophy*, pp. 104-5).

†See G. Gale in Leslie.

is the interesting one; it might be realized if, for example, it should turn out that the mechanism of the Big Crunch is such that it acts like an efficient document-shredder, totally destroying all information ("memory") of what happened before it. In such a case, it is presumably just a matter of linguistic taste whether we say that the events that "follow" the Crunch belong to a universe different from ours, or that our own universe is "reborn" from scratch. Clearly a similar analysis can be applied to possible events taking place "before" the Big Bang. However, in case (2), it is not clear that the words "before" and "after" have much real meaning.

A rather similar analysis applies to the possibility of universes separated spatially from ours. If there is any possibility of them communicating with us, either now or in the future (or past), then of course by our definition they are not "different," but are part of our own universe. If this is not to be the case, and we assume that our own universe is constructed in the usual geometry of general relativity (with 1 time and 3 space dimensions), then to make sense of the statement that the "other" universes are *spatially* separated from ours, we would need, *as a miminum*, to assume that their spatial geometry, at least, is the same as ours (in 3D), and moreover, that both ours and theirs is closed (though they could perhaps conceivably have more than one "time" dimension). Thus, we have as it were two "balls" in (or rather *of*!) 3D space that are guaranteed never to make contact. In a Newtonian picture, this may make sense: both "balls" can be visualized as floating in a sort of pseudo-3D – perhaps Euclidean – "background" space (though it is not clear what this means!), or perhaps more reasonably as embedded in a space of higher dimension (cf. the surfaces of two footballs, embedded in ordinary 3D space). In a Leibnizian picture, on the other hand, the whole concept of "spatial separation" of two closed universes would seem difficult to assign meaning.

Finally, we turn briefly to the possibility that there are other universes that are separated from our own neither spatially nor temporally, but in some other way. One obvious possibility is that they, as it were, occupy a different set of dimensions from ours (which would imply that "space" is more than 3+1–dimensional). To get some kind of intuitive feeling for what this possibility would look like, imagine that we take our familiar 3+1–dimensional Minkowski space with the space part described in spherical polar coordinates relative to a given origin, and cut two 2D subspaces out of it: one is simply the surface of a sphere of radius r_0 at some time t_0, and the other is defined by a spatial coordinate extending from some lower limit $r_1 > r_0$ to ∞ (with some fixed value of the angular coordinates) and the usual "time" coordinate. It is clear that there is no point of the original 4D (3+1) space that is common to the two subspaces, so assuming that the "dynamics" of the "universes" associated with a given subspace is restricted to that space, they will never interact, and by our criterion indeed constitute different "universes." In the same way, if there exist in some sense more than 4 dimensions, we might imagine one or more "parallel" universes existing in these extra dimensions with which we can, in principle, never communicate.

Probably the most-discussed example of universes that are "parallel" to ours, but not spatially or temporally separated from it, is, of course, the "many universes" of the Everett-Wheeler (relative-state) interpretation of QM. I will discuss this under the heading of "quantum cosmology", but in the present context remark that if the "different universes" into which the splitting is said in (the more imaginative versions of) this interpretation to occur are to be regarded as genuinely "different" according to our criterion, then it seems to be essentially required that the corresponding branches of the wave function are absolutely guaranteed never subsequently to show any interference. This is, in fact, usually claimed by adherents of the MWI to be true, at least once one reaches the level of human observation; but as in other "solutions" of the QMP, this raises the question of whether there is a definite point in the chain of macroscopic amplification at which one suddenly goes from a situation where interference can in principle occur, to one where it definitely cannot. How many "different" universes there are would seem to depend crucially on the answer to this question!

"Quantum cosmology"

I will deal only briefly with this supposed area of physics, because I personally regard it as a non-subject. Many people, and in particular many theoretical cosmologists (Hawking, Hartle, Carter...) are so impressed by the successes of QM as demonstrated in existing experiments (of the level of which they often seem to have only a rather hazy appreciation) that they assume almost without argument that it can be extrapolated to the behavior of the Universe as a whole, and in particular to the behavior of the scale factor $R(t)$. Formally, one expects that as with other macroscopic variables (except in very special circumstances; see end of lecture 23), the results of a QM calculation will not usually be very different from that of a classical one; the exception is the very early stages of evolution of the HBB, where R is $\leq$ the Planck length. However, there is a severe interpretational problem. QM as we know it was developed to deal with *ensembles* (of atoms, electrons, etc.), and the "standard" interpretation of the probability amplitude (wave function) is that it describes only such ensembles, not the individual members of them. But we have, by definition, only one Universe (or at least only one that we can ever know about); in this situation how can the concept of "ensemble" make sense? It was precisely this line of reasoning that led Everett (a cosmologist)[‡] to develop the MWI, according to which the wave function describes *individual systems* (in this case, the Universe as a whole!) and is never reduced in the act of measurement. This idea (and more generally the idea of the "wave function of the Universe") has been seized on by many subsequent theorists and used to try to do concrete calculations; in particular, it is sometimes supposed that it was quantum fluctuations in the "primordial soup" of the HBB that led to the formation of the nuclei of galaxies, and thus gave rise to the irregular pattern of galactic structures we see around us today. As with the more general case, there is an obvious question – at exactly what point any Everett "branching" took place so as to, as it were, stick us with the particular set of galactic structures we actually seem to have, rather than with any of the numerous (in fact infinite!) other possibilities – and cosmologists do not seem to be any better than other adherents of the MWI in answering this question. (As to the actual predictive success of this idea,

[‡]See *Physics World*, Nov. 2010, p. 36.

according to the best of my knowledge, it is at present neither better nor worse than various alternative models of the genesis of galactic structure – which is to say that its successes are at best qualitative and arguable.)

One very amusing result comes out of the basic notions of quantum cosmology, irrespective of whether or not one believes the MWI: if one takes the idea to its logical conclusion, it seems at first sight perfectly consistent to suppose that the Universe arose quite literally out of "nothing" by a giant quantum fluctuation! (Tryon: "In answer to the question of why it [the Big Bang] happened, I offer the modest proposal that our Universe is simply one of those things that happen from time to time.") But can we really make sense of the idea of a "period of time" before the Big Bang when there was "nothing"?

Before leaving this topic, I should mention that it is not clear that to embrace the idea of a "wave function of the Universe" necessarily commits one to embracing the MWI of quantum mechanics. In the last few years, there has been some very interesting work along the general lines of the Bohm–de Broglie "pilot wave" interpretation (note: not "theory"!), but now assuming that the wave function in question really does describe the Universe as a whole. (Recall that in the BdB theory, the wave function is a property of individual systems, not of ensembles.) It is a matter of considerable debate whether the metaphysical picture one gets out of this approach is really different from that apparently imposed by the MWI.

The anthropic principle

It is, of course, rather trivially true that one experimental constraint we have on models of our Universe is that it should permit, inter alia, the existence of human life. What is less trivial is that according to our current understanding of astrophysics, chemistry and biology, the physical conditions for the existence of *any* kind of life, and hence, a fortiori, of the human variety, are so delicate that, apparently, even a very small variation of any one of a number of parameters away from their existing values would be sufficient to destroy them. For example, the biochemical reactions essential to life depend extremely sensitively on the energies of the molecular states involved, which in turn are very sensitive to the exact value of the electron mass and charge; were the electron charge only very slightly different from what it in fact is, the biochemistry necessary to support life as we know it could not exist. Again, the development of life apparently requires not only the right chemical conditions but just the right distribution of incident radiation; were the ratio of electromagnetic and gravitational interaction constants only slightly different from what it actually is, our sun would be unable to provide the correct mix. In the context of recent grand unified theories, which allow the proton to decay, one observes that a small variation in the fundamental constants would allow a proton lifetime much shorter than the current lower limit of $\sim 10^{33}$ yrs, in which case we would all be dying of radiation sickness. Again, in a world that had other than three spatial dimensions, the gravitational force would not follow an inverse-square law, hence stars as we know them could not exist, hence no planets… and so on. The list could be multiplied almost endlessly. It is then easy to draw the conclusion that for any kind of conscious beings to exist, the basic physical constants have to be exactly what they are, or at least very close to them. What makes these considerations particularly suggestive is that in fact the values of the various "critical" parameters (electron-proton mass ratio, dimensionless electron charge, cosmological density parameter, dimensionality of spacetime. . .) are not themselves fixed by any known considerations in our current particle physics or cosmology.

It is considerations of this kind that lead to the formulation of the famous (or infamous!) *anthropic principle*, which in its most generic (and most ambiguous!) form says, in effect: the reason that the Universe is the way it is is because human beings are here to observe it. It should be noted right away that when unpacked, this claim can be interpreted at very different degrees of strength. At its weakest, it is simply a summary of the arguments illustrated above, to the effect that if the constants of nature were only very slightly different from what they are, we should not be here to wonder about them. At its strongest, it makes the essentially teleological claim that the *reason* the physical constants have the values they do is so as to permit the existence of human life. It is interesting to note that even this strong form of the anthropic principle actually has a very long history, predating modern physics as we know it; and even within the context of the latter, it goes back at least to Darwin's collaborator Alfred Wallace. It is also interesting that the first self-conscious formulation in the context of modern cosmology, by Brandon Carter in 1973, was expressed quite explicitly as a reaction against what he regarded as an excessive reliance on the "Copernican principle" that the situation of man in the Universe is in no way special.

At its weakest, the anthropic principle is simply a list of coincidences that are not necessarily in themselves of any great philosophical interest (and, incidentally, it is easy to exaggerate their significance: for example, if in fact the radioactivity due to proton decay were much larger than it is, it is not at all obvious that nature could not have evolved alternative forms of organism to cope with this; nor do all biochemists agree that the only viable life forms must necessarily be carbon-based). At its strongest, the maintenance of the principle would seem natural only within an explicit teleological framework, e.g., one based on the idea of divine providence – a kind of framework that most physical scientists today, irrespective of their religious beliefs, regard, rightly or wrongly, as inappropriate in the context of the explanation of physical laws. The most interesting, and philosophically most controversial, uses of the principle lie somewhere between these extremes. Let's take, for example, the question: why is the age of the Universe what it is (i.e., somewhere between 7 and 15 billion years, according to the best current estimates)? To what extent is it a valid answer to this question to point out that this is precisely the kind of period necessary, according to our current notions in astrophysics, geology, etc., for stars to form, planets to congeal out of the stellar gas, planetary temperatures to cool to the point where life can develop, etc.? To the extent that we regard this answer as valid, we are in some sense rejecting the extreme application of the "Copernican principle", which would say that not only is there no good reason to suppose that we live at a special place in the Universe, there is also no good reason that we should live at a special time in its history.[§]

A more interesting application, though, is to the question of the values of the fundamental constants (and the dimensionality of spacetime, etc.). Can one make sense of an "anthropic" explanation in this context without committing oneself to unwanted teleological baggage? In the context of a single unique (and homogeneous) Universe, this seems difficult. However, within a "many universe" scenario, it might make some sense – or for that matter, in one in which, within a single Universe, parts that are widely spatially separated behave in radically different ways. For example, in some scenarios of so-called "chaotic inflation", the original high degree of symmetry in the very early Universe is "broken" in different ways in different regions that at the time were causally disconnected. If the Universe is in fact spatially infinite, there could then evolve an infinity of different regions, each with its own values of the "fundamental constants" such as mass, charge, etc. The fact that we humans inhabit a region where the constants are what we observe them to be is then in some sense not an accident – in most other regions, we could not have evolved. Of course, within this particular scenario one might expect that in the distant future we might come into contact with those "other" regions (this possibility is part of the definition of a "single" universe!).

If the idea works to an extent in a single universe, it should a fortiori work in a "many-universe" scenario in which the other universes are, by construction, further hidden from our view. A particularly intriguing implementation is the "many-worlds" interpretation of QM: in this case there might be many "branches" of the universal wave function that correspond to conditions under which life could not have evolved, and the reason why we (seem to) find ourselves on one of the ones in which it could, is obvious!

Irrespective of the particular variant of the anthropic principle, there is one quite general point that is worth noting: however severely the requirement of compatibility with human existence constrains the fundamental constants, as far as we know today, the values of the latter that are allowed in the absence of the former form a *continuum*,[¶] and thus the anthropic considerations can at best constrain them to lie within a band of values of non-zero width.

Thus, prima facie, the anthropic principle can never be a *complete* explanation of the values found experimentally. This state of affairs might perhaps change if either (a) we should discover some fundamental reason why quantities like the electron charge and mass can take only discrete values, or (b) we should discover some fundamental limitations on the accuracy with which we measure such quantities. However, at present, no such limitations are known to exist.

[§] If, contrary to our current beliefs, our Universe were to turn out to be strongly *spatially* inhomogeneous, a similar question might be raised as to whether it is an accident that (e.g.) we live nearer the center than the "edge".

[¶] The situation is clearly different as regards (e.g.) the dimensionality of spacetime or the cosmological parameter k, which can take only discrete values.

Irreversibility and the Big Crunch

If anything, questions concerning the arrow of time in a cosmological context are even more "slippery" than in a general one. To quote Price,

> cosmologists who discuss these issues often make mistakes which are strikingly reminiscent of those which plagued the nineteenth-century discussions of the statistical foundations of thermodynamics. The most common mistake is to fail to recognize that certain crucial arguments are blind to temporal direction, so that any conclusion they yield with respect to one temporal direction must apply with equal force with respect to the other... The fundamental lesson of those endeavors is that much of what needs to be explained about temporal asymmetry is so commonplace as to go almost unnoticed. In this area more than most, folk intuition is a very poor guide to explanatory priority.

As we have seen, a plausible, if not certain, view is that the other "arrows" of time (excepting the elementary-particle one) can be regarded as derivative from the cosmological one. In the context of the latter, there are really two different, but related, points at issue. One, which can be raised irrespective of the value of Ω_0 (i.e., of whether the Universe is open, flat or closed), is: why was entropy so low at the Big Bang? The second, which is relevant only in a closed (and hence prima facie time-symmetric) Universe, is: what will happen at the Big Crunch? In particular, will the arrow of time reverse as the Universe goes through the point of maximum expansion and starts to contract? I will explore these questions in turn.

We said earlier (lecture 25) that the thermodynamic and other arrows of time could be a result of the fact that stars are emitting rather than sucking in radiation, which in turn is associated with the fact that the Universe is expanding rather than contracting. But it should be strongly emphasized that the nature of this latter "association" is not at all clear. One cannot (at least on the basis of our current understanding) assert that in *any* arbitrary expanding universe, stars would be formed in the way they are in ours. Indeed, according to our current picture (which of course assumes, inter alia, that we can extrapolate ideas about gravity, etc., to the Universe on a large scale), in order for galaxies and eventually stars to be formed in the way we see them, the Universe has to have started out, at the Big Bang, in a state that was very nearly, but not quite, completely homogeneous. (If it had been more homogeneous, then [at least in classical physics]‖ galaxies would never have formed; if less, nearly all the matter would have been gobbled up into giant black holes at an early stage.) Now one might think at first sight that such a nearly completely homogeneous Universe would correspond to a state of very high "randomness" and therefore very high entropy. But this is not so: the state is too homogeneous to be random! It is as if one were to examine a long sequence of coin tosses by splitting it up into groups of 100 successive throws, and to find that in each group of 100 there were *exactly* 50 heads and 50 tails – this is actually an exceedingly *improbable*

‖Cf., however, the remarks above about nucleation by quantum fluctuations.

configuration. In the case of the Universe, it has been estimated (Penrose) that the fraction of possible configurations having the required very high degree of "smoothness" is around 1 part in $10^{10^{120}}$! To be sure, the inflationary scenario helps a little here, but by no means enough. Of course, once given this very improbable starting situation, the entropy of the Universe as a whole cannot help but increase – but we note that this is not (or at least not obviously) a consequence of expansion *per se*.

So why is it that the Universe started off in this state of very low, but nonzero, entropy? Of course, one possible reply is that the whole concept of entropy really makes little sense when applied on this scale – perhaps because we really have no justification for the implicit application of the principle of indifference to the Universe as a whole. Maybe the question should not even be asked; i.e., we should take the "very improbable" initial condition as a fact of life, on the same footing as (say) the value of the dimensionless electron charge. Alternatively, we could perhaps invoke some version of the anthropic principle – the initial conditions on "our" Universe are indeed statistically very unlikely, but had they been different, we would not have been here to ask the question. A third line of solution is the so-called "no-boundary condition" of Hawking and Hartle; however, this involves the essential use of so-called "imaginary time", and to many people appears a mere mathematical trick with no real physical context. So the question remains very much an open one.

Let's now suppose that we have been able, in one way or another, to justify the postulate of low initial entropy, i.e., entropy that decreases backwards in time towards the initial singularity. What we have then shown is that entropy tends to increase as the Universe expands, and in an open or flat Universe that is more or less the end of the matter (such a Universe will go on expanding forever, and the entropy will continue to increase, until presumably all matter eventually ends up in the form of various black holes). However, in a closed Universe, the urgent problem arises of the behavior of the entropy in the contracting phase. As forcefully emphasized by Huw Price, the problem is that any argument that says that the entropy was very low at the Big Bang will prima facie also lead to the conclusion that it will be small at the Big Crunch. Thus, such arguments would say that entropy *decreases* in the contracting phase and hence that the arrow(s) of time will be reversed in that phase. Such a scenario is sometimes called a "Gold universe". Although there are some intriguing problems concerning what would actually be "felt" by an observer who happened to live through the point of maximum expansion, it does not seem there there are any fatal a priori objections to this scenario.

However, it appears that most contemporary cosmologists do not like this conclusion and prefer to assert that the current thermodynamic, and hence psychological, arrow of time would remain valid in the contracting phrase, so that the entropy at the Big Crunch is enormous. In particular, this appears to be the current view of Hawking (*Brief History of Time*, p. 150). But assuming one believes (as most people other than Penrose seem to) that the microscopic laws of physics are time-symmetric, most, if not all, of the arguments given to this conclusion indeed seem to involve a "double standard" – the failure to recognize that an argument that applies in one direction in time will equally apply in the other. (This remark includes the Hawking-Hartle NBC idea along with the others.)

The only way out seems to be the following. Suppose we could prove that although the basic laws of physics are themselves time-symmetric, the majority of relevant solutions "break" this symmetry; i.e., with a "direction" of time specified a priori, they involve *either* a strong increase *or* a strong decrease of entropy with time. This is not a priori totally unreasonable: compare the fact that in a magnetic material at low temperature, although the basic interactions are invariant under space reversal, the thermodynamic equilibrium state either has a majority of spins pointing up or has a majority pointing down, the two macroscopic states being equally probable. Of course, just as in that case, the a priori probability of an entropy-increasing solution would be identical to that of an entropy-decreasing one. Thus, it is guaranteed with high probability that the Universe we live in, though symmetric in time with respect to its expansion and contraction, is nevertheless asymmetric in that the entropy either increases in both expansion and contraction phases, or decreases in both (with, as noted above, some arbitrary a priori convention chosen for the "direction" of time). At this point, we may legitimately invoke a fairly weak version of the anthropic principle: if according to the arbitrary convention, entropy in our Universe is actually decreasing, we will nevertheless "experience" it backwards; i.e., our psychological arrow (and our identification of the thermodynamic arrow, etc.) will be *opposite* to that of the convention! So whether or not entropy is "really" increasing or decreasing, we will always "see" it decreasing.

The only problem is: to the best of my knowledge, there exists no proof, or even plausible argument, for the above conclusion! It is clear that here we stand right at the edge of our understanding...

[If time permits in lecture, we will discuss the possibility of spatially and temporally limited "fluctuations in the direction of time".]