General Relativity and the Eternal Return*

by

Frank J. Tipler
Department of Mathematics
University of California at Berkeley
Berkeley, California 94720
U.S.A.

Abstract

An arbitrarily close return to a previous initial state of the Universe, such as is predicted by the Poincaré recurrence theorem, cannot occur in a closed universe governed by general relativity. The significance of this result for cosmology and thermodynamics is pointed out.

"The thing that hath been, it is that which shall be; and that which is done is that which shall be done: and there is no new thing under the sun."

—Ecclesiastes 1:9

"God forbid that we should believe this. For Christ died once for our sins, and rising again, dies no more."

—St. Augustine (City of God, XII, ch. 13)

*Work supported by the National Science Foundation under grant number PHY-77-15191

The idea that history repeats itself - that every state of the Universe has occurred before and will occur again __ad infinitum__ - is a concept which originated thousands of years ago and in the form of the Poincaré cycle and the "cyclic" universe, still plays an important role in fundamental physics. This notion of an "eternal return" will be studied in this paper in the context of general relativity, and it will be shown that in a closed universe, no arbitrarily close (or exact) return to a previous state of the universe is possible, provided certain very general restrictions on the matter tensor and global causal structure hold. This result is in stark contrast to the situation in classical mechanics: Poincaré showed [1,2] that for almost all initial states, any mechanical system with a finite number of degrees of freedom, finite energy, and constrained to move within a finite box must necessarily return arbitrarily closely and infinitely often to almost every previous state of the system. The reason for "no return" in general relativity and "eternal return" in classical mechanics is that in general relativity, singularities intervene to prevent recurrence. General relativistic closed universes are thought to begin and end in singularities of infinite curvature, and these singularities force time in general relativity to be linear rather than cyclic.

The notions of recurrence vs no-return constitute what Holton [3,4] calls a thema-antithema pair. Briefly, these are pairs of opposite fundamental concepts which form the basic framework of all scientific models. Other examples are absolute vs. relative, the plenum vs. the void, atomism vs. the continuum, and determinism vs. indeterminism. Holton argues that although scientific theories change, the fundamental thema (whose number he estimates [4] to be less than 100) remain in different guises and different proportions in suceeding theories. This is certainly true of the recurrence/no-return couple. A brief history of the eternal return concept in philosophy and science will be given in section 2, with emphasis on the versions of this idea which have strong parallels in modern science. Section 3 will give a precise statement of the general relativistic no-return theorem, together with a brief explanation of the meaning of the terms used in the theorem. The proof of the theorem will be given in section 4.

As pointed out by numerous authors [2,5,6,7] Poincaré recurrence is a major stumbling block to the definition of entropy, a function of the state of the system which never decreases, in terms of the fundamental microscopic variables of the system. Hitherto the increase of an "entropy" defined by such variables has been obtained (not very successfully) by such tricks as "coarse-graining" [2,8]; an ad hoc addition of randomness to the system in the form of a postulate of "molecular chaos" or "random phases";[2,5] or by taking the thermodynamic limit. [9,10] It will be argued in section 5 that the results of this paper may make such subterfuges unnecessary, and that there is a deep connection between the increase of gravitational field entropy - which does not seem to require the above tricks - and the increase of matter entropy. The paper will conclude with a discussion of the significance of the no-return theorem for cosmology.

Brief History of the Eternal Return Idea

The concept of the eternal return - the idea that time is fundamentally cyclic - apparently played a key role in the cosmological thought of mankind as far back as 6,500 B.C. ([11], p. 332). Scientific thought in this period was based on such common sense phenomena as the cycle of the seasons, the rhythm of human life from birth to adulthood to death, and the numerous periodicities in the heavens such as the phases of the moon, the annual motion of the sun through the constellations, and the periodic near return of the planets (Gods: see [11], p. 4) to previous positions in the sky. Under these circumstances a cyclic notion of time is more natural than rectilinear time, and it was cyclic time that dominated the thought of the so-called primitive peoples.[12,13,14]

The early agricultural civilizations - Sumer, Babylon, [12], p. 44; [15], p.360) Indian,([15], p. 353), Mayan, ([13], p. 88), and Shang-Chou, ([15], p. 358) - retained and elaborated the notion of cyclic time. The Babylonians, for example, based their concept of time on the periodicities of the planets. In their view, the lifetime of the universe, or Great Year, lasts about 432,000 years. The summer of this Great Year would be marked by the conjunction of all the planets in Cancer, and would be accompanied by a universal conflagration; the winter would occur when all planets have a conjunction in Capricorn, and this would result in a universal flood. The cycle then repeats and in some accounts the next cycle is an exact reproduction of the preceding ones. ([15], p. 362). The ancient Indians (Hindus, Buddhists, Jains) extended this basic structure of a single Great Year into an entire hierarchy of Great Years. For instance, a destruction and re-creation of individual forms and creatures (but not of the basic substance of the world) occurred every Kalpa or day of Brahma. Each day of Brahma had a duration of about 4 billion years. The Elements themselves together with all forms undergo a dissolution into Pure Spirit which then incarnates itself back into matter every lifetime of Brahma, or about 311×10^{12} years. ([15], p.363; [16], p. 354). The Brahma lifetime is the longest cycle in the Indian system, and the cycle is repeated ad infinitum.

Among the Greeks, the Stoics were the most fervent believers in the eternal return. They held ([14], p. 47) that all objects in the universe were bound together in an absolutely determinate web of actions and reactions, and that this determinism led to a precise recurrence of all events. That is, no event is unique and occurs once and for all (for example the condemnation and death of Socrates) but rather every event has occurred, occurs, and will occur, perpetually; the same individuals have appeared, appear, and reappear in every return of the cycle. Cosmic duration is thus repetition and anakuklosis, eternal return. ([13], p. 89)

This Stoic idea of palingenesia - that is, the reappearance of the same men in each cycle ([12], p. 47), carried the idea of the eternal return to its logical extreme, and went much further than the above-mentioned earlier thinkers were willing to go, or indeed Aristotle and Plato were willing to go.

Aristotle was intrigued by the notion of palingenesia. He noted that if it were true it would obscure the usual idea of before and after, for it would imply ([17]; [14], p. 46) that he himself was living as much before the fall of Troy as after, since the Trojan War would be re-enacted and Troy would fall again. However, although he accepted the cycles, he was reluctant to accept the exact identity of events in each cycle, arguing that the identity was only one of a kind ([12], p. 48). Plato's cosmology was also cyclic, with a periodic destruction and re-creation of the universe in conjunction with various astronomical events [18,19]. Indeed, it is through Plato's writings that the notion of the Great Year entered later Western thought. However, scholars disagree as to whether his concept of the cycles went to the extreme of palingenesia ([12], p. 48; [14], p. 45). The eternal return concept came to dominate thought in the pre-Christian, later Roman empire. It was also prominent on the other side of the oikoumene in this period, namely in Han China. As Needham has pointed out ([20], p. 29), the popular religious Taoism of the Han was millenniarist and apocalyptic; the Great Peace was clearly in the future as well as the past. In the Canon of the Great Peace (written between 400 B.C. and 200 A.D.) is found a theory of cycles which issued fresh from chaos and then fell slowly until the day of doom.

As the epigram to this paper suggests, the Christian world view was hostile to the eternal return idea; in the section of the City of God from which the selection quoted is taken, St. Augustine was criticizing the Stoic recurrence concept, arguing that Christian philosophy (and its Hebrew predecessor) required a non-cyclic linear concept of Time. God created the world once, Christ died once, and will rise again once. With the triumph of Christianity, the notion of linear time became dominent over cyclic time in the West until the rise of modern science, though a few Medieval scholars, such as Bartholomaeus Anglicus (1230), Siger of Bruhant (1270) and Pietro d'Acono (1300), were willing to at least entertain the notion of an eternal return. [21] In Medieval China, however, the Neo-Confucian school, which flourished in the 11th through 13th centuries of our era and which was influenced both by Buddhist ideas on recurrence and the above-mentioned ideas of ancient Taoism, accepted the idea that the universe passed through alternating cycles of construction and dissolution.([20], pp.6,22). For instance, the Sung scholar Shen Kua (f. 1050) discussed recurrent world-catastrophes ([20], p. 22; [22], pp. 598 ff and 603 ff), and later the Ming Scholar Tung Ku held that a world-period had a beginning, but not the endless chain of all world periods ([20], p. 6; [22], p. 406). Needham has argued ([20], p. 50) that in fact the linear notion of time dominated over the cyclic view in later Chinese thought, but other scholars disagree.[23] However, there is no question that linearity dominated in Christian thought, and many scholars (e.g. [16,24]) have contended that this notion of time played a key role in the rise of modern science.

Modern science in turn led to a revival of cyclic time. The Newtonian world picture contained both cyclic and linear aspects from the very beginning. Newton himself was worried that his solar system model - based on a linear (mathematical) time - was gravitationally unstable in the long run, and to compensate for this instability he suggested a cyclic process whereby the planets would be replaced as they were periodically perturbed from their orbits by the gravitational action of other bodies.[25] Euler, Laplace, Lagrange and others had shown by the beginning of the 19th century that the solar system was in fact stable to first order, the gravitational perturbations leading merely to a cyclic oscillation of the planetary orbits. However, at about this time the debate on the question of cyclic vs. linear shifted from astronomy to geology and thermodynamics. ([5], p. 553). The problem in geology was whether the internal heat of the earth could power geological cycles indefinitely, or whether the earth would eventually cool to a "state of Ice and Death", as J. Murray[26] put it in 1814. This question in part stimulated research in thermodynamics ([5], section 14); by the end of the 19th century, Kelvin and others (see [5] for a detailed history) concluded on the basis of the newly formulated Second Law of Thermodynamics that the Heat Death was inevitable, and hence a cyclic notion of time was refuted. Kelvin[27] and Tait[28] indeed inferred that the Second Law implied a creation of the universe.

Other physicists were unwilling to grant such unlimited validity to the Second Law, arguing that the creation of the Universe would violate the First Law of Thermodynamics.[29] Thus somehow the energy dissipated by thermodynamics processes must be periodically reconcentrated into usable from. Rankine, for instance, suggested[30] that heat radiated into space would reach a sort of "ether wall" a finite distance from the earth, at which the radiant heat would be totally reflected and reconcentrated into various "foci". (Rankine's notion of "ether wall" is strikingly similar to the idea of "domain boundary" which arises in spontaneous symmetry breaking in gauge theories [31,32]). Thus the history of the Universe would be cyclic in the long term.

Rankine was basically trying to show that mechanics and the Second Law were inconsistent. This was first shown in 1890 by Poincaré in his famous recurrence theorem [33] mentioned above. In its most general form Poincaré's theorem can be proven in any space X on which there is a one parameter map T_t from the sets $\{U\}$ of X into $\{U\}$ and a measure μ on X such that: (1), $\mu(X) = 1$ and (2), $\mu(T_{t_0}(U)) = \mu(T_{t_0+t}(U))$ for any $U \subset X$ and any t_0, t. In the application to classical mechanics, (1) is assured by requiring the space X to be the phase space of a finite energy mechanical system in a finite box. If μ is the density function ρ in phase space and T_t is the evolution operator for the mechanical system (which is assumed to be Hamiltonian), (2) then follows from Liouville's theorem: $d\rho/dt = 0$. Thus the classical mechanics of a finite system <u>is</u> inconsistent with the Second Law; by Poincaré's theorem, almost all such systems must return arbitrarily closely and infinitely often to almost all previous initial states.

At about the same time Poincaré was developing his theorem, the English philosopher H. Spencer [34] and the German philosopher F. Nietzsche were attempting to make a scientific-sounding argument for the eternal return. Nietzsche's argument is worth repeating in detail because although it is non-rigorous (to say the least!) it contains all the essential ideas which are needed for a rigorous proof (given certain assumptions about the evolution of the world-system), a proof which will be useful in discussing the significance of the Theorem stated in the next section. I shall follow the example of Nietzsche's sister and omit those parts of his argument which I think are nonsense.

Nietzsche began his "proof" of eternal recurrence as follows:

> ...we insist upon the fact that the world as a sum of energy must not be regarded as unlimited - we forbid ourselves the concept infinite energy, because it seems incompatible with the concept energy ([35], #5).

This is similar to the idea of energy in general relativity. Only in asymptotically flat space where the total energy is necessarily finite does the concept energy have a well-defined meaning ([37], p. 457). Nietzsche also argued that the universe must be infinite in time:

> We need not concern ourselves for one instant with the hypothesis of a created world. The concept "create" is today utterly indefinable and unrealisable; it is but a word which hails from superstitious ages... ([36], #1066).

He then contended that recurrence of all states followed from the finiteness of energy (and space) by which he meant a finite number of possible states of the universe, the infinity of elapsed time, and a chance-like evolution:

> If the universe may be conceived as a definite quantity of energy, as a definite number of centers of energy, -

and every other concept remains indefinite and therefore useless, - it follows therefrom that the universe must go through a calculable number of combinations in the great game of chance which constitutes its existence. In infinity, at some moment or other, every possible combination must once have been realized; not only this, but it must have been realized an infinite number of times...([36], #1066). If all possible combinations and relations of forces had not already been exhausted, then an infinity would not yet lie behind us. Now since infinite time must be assumed, no fresh possibility can exist and everything must have appeared already, and moreover an infinite number of times. ([35], #7). That a state of equilibrium has never been reached, proves that it is impossible. [But it must have been reached] in spherical space ([36], #1064). Only when we falsely assume that space is unlimited, and that therefore energy gradually becomes dissipated, can the final state be an unproductive and lifeless one. ([35], #8).

In the last section of this paper, it will be pointed out that Nietzsche's world-model can be compared fairly closely to a Markov process whose states must recur. Thus, granting Nietzsche's assumptions (finite number of states, no creation and chance-like evolution), his proof of recurrence is valid when judged by the standards of philosophical (not mathematical) rigor.

Another 19th century thinker who considered the problem of recurrence was Boltzmann. Boltzmann originally hoped to deduce irreversibility from the mechanics of atoms but he soon realized such a deduction was impossible without using averaging techniques. Under pressure from Planck's student Zermelo, who based his arguments on Poincaré's theorem, Boltzmann suggested that the Universe as a whole had no time direction, but individual regions of it did, when by chance a large fluctuation from equilibrium should produce a region with reduced entropy. These reduced entropy regions would then evolve back to the more probable state of maximum entropy, and the process would repeat in accordance with Poincaré's theorem. ([38, 5], section 14.7).

Once it became clear that a finite system of particles would be recurrent and not irreversible in the long run, Planck considered whether irreversibility could arise from a field theory such as electromagnetism. The idea would be to derive irreversibility from the interaction of a continuous field with the discrete particles. Planck began a series of papers on this question in 1897, a series which culminated in his discovery of the quantum theory of radiation in 1900. Boltzmann, however, pointed out that if we regard the field as a system with an infinite number of degrees of freedom, this would be analogous to a mechanical system with an infinite number of molecules, and in either case we would have an infinite Poincaré recurrence time; in either case we would have irreversibility and long term agreement with the Second Law. However, for the thermodynamics of fields in a confined space, it is physically more appropriate to regard the field not as a continuous quality governed by differential equations, but rather as a large but finite number of "vector ether atoms" whose equations of motion are obtained by replacing the usual differential equations with finite difference equations. To this system the recurrence theorem would apply ([39], [5], section 14.8).

Most Twentieth century discussions of the eternal return are based on the so-called oscillating closed universe model developed in 1922 by A. Friedmann.[40] Friedmann himself was aware of the cyclic nature of time in his solution, and suggested that one could identify corresponding times in each cycle. However, in the Friedmann model the radius of the Universe went to zero at the beginning and the end of each cycle, and thus from a strict mathematical standpoint the cycles were disjoined by a singularity; they were not true "cycles." Tolman proved[41] in 1931 that such a discontinuity was inevitable at the beginning and end of any isotropic and homogeneous closed universe with a physically reasonable matter tensor. He argued[42] that this discontinuity was merely an artifact of the high symmetry assumed, and in a physically realist universe, the actual discontinuity would disappear. He therefore assumed that the entropy would be conserved in a passage through the singularity, and thus the thermodynamics of a cycle would in part be determined by the history of a previous cycle. Other relativists of the time by and large agreed with Tolman as to the unreality of the singularity - see ref.[43] for more details.

With the advent of the Hawking - Penrose singularity theorems in the 1960's, most relativists have accepted the reality of some sort of initial singularity - at least in classical relativity. Some relativists have argued that quantum effects could cause the universe to "bounce" at very high densities, leading to cycles in the closed universe models. Wheeler, for example, has until recently suggested [44] the physical constants themselves are recycled at a "bounce." Thus Tolman's concept of the cycles was analogous to the "Day of Brahma" cycle in Indian mythology, while Wheeler's resembled the "life of Brahma" cycle. I shall now show that in classical relativity, the singularity prevents recurrence, and in Section V I shall argue that if quantum effects do result in a "bounce", that some sort of recurrence is probably inevitable.

III. The No-Return Theorem

In order to prove that no two states of the Universe can be identical or even arbitrarily close, we must first make precise the notion of "close." This can be done by regarding the set of all initial data as a Sobolev space W^s. The global topology of the spacetime (M,g) is $S \times \mathbb{R}$, where S is compact, if the spacetime is globally hyperbolic. We choose some positive definite metric e_{ab} on M, and define a norm on W^s by

$$\| K^I{}_J \|_m \equiv \left[\sum_{p=0}^{m} \int_S (| D^p K^I{}_J |^2 d\sigma \right]^{\frac{1}{2}} , \quad (1)$$

$$\| h, \chi \| \equiv \| h \| + \| \chi \|$$

where $d\sigma$ is the volume element induced on S by e_{ab}, D^p is the generalized p^{th} covariant derivative with respect to some chosen background metric $\bar{g}_{ab}$, and $| \ |$ is the norm induced by e_{ab} (see pp. 233-235 of [45] for more details). Two tensors will be "close" if they are "close" in the norm (1).

The No-return theorem will require four conditions on the matter tensor. The first condition, the timelike convergence condition ([45], p. 95), says roughly that gravitation is always an attractive force. The other three conditions, which will here be called (a), (b), and (c), are precisely stated on pages 254-255 of [45]. Roughly stated, (a) says that the development from a given set of initial data is locally unique; (b) says that this unique development is locally stable; and (c) says that the stress-energy tensor is a polynomial in the matter fields, their first derivatives, and the spacetime metric.

As discussed on page 255 of [45], these are physically reasonable conditions to impose on the matter fields, though strictly speaking, they are not necessary conditions. Any other conditions which imply global Cauchy stability and uniqueness (in the sense defined in Chapter 7 of [45]) would be sufficient to prove the Theorem below.

We shall also need two global conditions. The first, the requirement of unique development from Cauchy surfaces ([45], p. 205), says that Laplacean determinism holds on the spacetime. (It is interesting that in general relativity, determinism implies no-recurrence, whereas in the Stoic philosophy and classical mechanics, determinism was held to imply recurrence.) The second, the generic condition ([45], p. 101) says that every causal geodesic feels a tidal forces at least once in its history. As this would be expected to be false only for a "measure-zero" set of solutions to the Einstein equations, the generic condition in the Theorem below is somewhat analogous to the "measure-zero" qualifications in the Poincaré recurrence theorem. A closed universe is defined to be a spacetime in which the Cauchy surfaces are compact. A spacetime that contains two disjoint spacelike Cauchy surfaces which are isometric in their initial data is said to be time periodic.

Theorem: If a spacetime (M,g) containing compact Cauchy surfaces is uniquely developed from initial data on any of its Cauchy surfaces, and if (M,g) also satisfies both the generic and the timelike convergence conditions, then the spacetime cannot be time periodic. Furthermore, if in addition the matter fields ψ and

<u>their first derivatives ψ' satisfy conditions (a), (b), and (c),</u> <u>then for any neighborhood U of any Cauchy surface</u> S_1 , <u>there exists a number $\epsilon > 0$ such that</u> $\|(h, \chi, \psi, \psi') - (h_1 , \chi_1 , \psi_1 , \psi_1')\|_{5 + a} > \epsilon$ <u>for the initial data on any Cauchy surface S with $U \cap S$ empty</u>. (ϵ depends on e_{ab}, $\bar{g}_{ab}$, and U . S and S_1 are assumed spacelike.)

V. Proof of the No-Return Theorem

Assume on the contrary that (M,g) is time periodic. By the assumption of unique development, this implies there exists a sequence of Cauchy surfaces $S(t_{-n}) \ldots S(t_0) \ldots S(t_n)$ with the initial data on a given surface isometric to the initial data on any other surface, and with $J^+(S(t_i)) \cap J^-(S(t_{i+1}))$ isometric to $J^+(S(t_j)) \cap J^-(S(t_{j+1}))$ for any i,j. Thus (M,g) can be regarded as a covering space for a spacetime with topology $S \times S^1$. By the corollary on page 217 of [45], there will exist a timelike curve γ_{ij} of maximal length between any two of the isometric Cauchy surfaces $S(t_i)$ and $S(t_j)$, and γ_{ij} will be orthogonal to both. Consider the sequence of curves $\gamma_{-1,1}$, $\gamma_{-2,2}$, ... $\gamma_{-n,n}$ .... Because $S(t_0)$ is a Cauchy surface, each of these curves intersects $S(t_0)$ in exactly one point. Since $S(t_0)$ is compact, the sequence has a subsequence γ_n which converges to a timelike geodesic γ. (Timelike since in the space $S \times S^1$, we can define γ by a sequence of vectors normal to $S(t_0)$ - the vector at the future endpoint of γ_n and tangent to γ_n there. This sequence of vectors has a subsequence which converges to a normal of $S(t_0)$, and all convergent subsequences converge to a normal to $S(t_0)$.)

The geodesic γ is past and future complete. To see this, we first show that the lengths of the geodesic segments γ_n must diverge in both the past and future directions as $n \to \infty$. For since γ_n is the maximal length geodesic segment between $S(t_{-n})$ and $S(t_n)$, and if γ_n converged to a finite length in either direction, the future direction say, then we could construct a causal curve between $S(t_{-n})$ and $S(t_n)$ of length greater than γ_n for n sufficiently large as follows. Every timelike geodesic normal to $S(t_0)$ reaches all $S(t_n)$ since these are Cauchy surfaces. Define a timelike curve $\alpha_n(p)$ from any point p in $S(t_0)$ by extending the geodesic normal to $S(t_0)$ at p until it reaches $S(t_1)$ at p_1, then move along the geodesic normal to $S(t_1)$ at p_1 until this geodesic reaches $S(t_2)$, and so on until $S(t_n)$ is reached. Since $S(t_0)$ is compact, the length of a geodesic from $S(t_0)$ to $S(t_1)$ along a geodesic normal to $S(t_0)$ is bounded below by some number L. Thus the length of $\alpha_n(p) \geq n\,L$, and so if the length of γ_n did not diverge in the future direction as $n \to \infty$, we could replace γ_n by $[\gamma_n \cap J^-(S(t_0))] \cup [\,\alpha_n(p = \{\gamma_n \cap S(t_0)\})\,]$ to obtain a causal curve of length greater than γ_n between $S(t_{-n})$ and $S(t_n)$ for n sufficiently large. But this is impossible, since by definition γ_n is the maximal length curve between these two Cauchy surfaces. By the continuity of length along arbitrarily close continuous geodesic segments, this divergence in both time directions of the lengths of the γ_n segments implies the geodesic completeness of γ.

Since γ is geodesically complete, and since the generic and the timelike convergence conditions hold, γ must have a pair of conjugate points - at points p and q, say - by Proposition 4.4.2 of [45]. By Proposition 7.24 of [46], the location of first conjugate points varies continuously with the geodesic, and so there will be points p_n, q_n on γ_n which are conjugate points on γ_n and converge to p, q. For n sufficiently large, p_n, q_n will be in $J^+(S(t_{-n})) \cap J^-(S(t_n))$. But γ_n is the maximal length causal curve between $S(t_{-n})$ and $S(t_n)$, and so by Proposition 4.5.8 of [45], it cannot have conjugate points in $J^+(S(t_{-n})) \cap J^-(S(t_n))$. (We could also obtain a contradiction by arguing as on page 270 of [45].) This contradiction shows that time periodic spacetimes do not exist.

We shall now show that it is not possible to approach arbitrarily closely to a previous initial state. The proof will be a generalization of an argument by Fischer and Marsden[47], who showed that if one gets sufficiently close to a previous initial data set of a Cauchy surface in empty space, then one can perform a coordinate transformation to obtain a time periodic spacetime.

If there is no such ϵ for some S with initial data (h, χ, ψ, ψ'), then there will exist a sequence of Cauchy surfaces S_n with initial data $(h_n, \chi_n, \psi_n, \psi'_n)$ such that $(h_n, \chi_n, \psi_n, \psi'_n) \to (\bar{h}, \bar{\chi}, \bar{\psi}, \bar{\psi}')$ as $n \to \infty$, with $S_n \cap U$ empty for all n, and with $(\bar{h}, \bar{\chi}, \bar{\psi}, \bar{\psi}') = (h, \chi, \psi, \psi')$. Let $(\bar{M}, \bar{g})$ be the maximal development of $(\bar{h}, \bar{\chi}, \bar{\psi}, \bar{\psi}')$. Now the topology of $\bar{M}$ and M is $S \times \mathbb{R}$, so we let $\bar{i}: S \to \bar{M}$ be an embedding inducing $(\bar{h}, \bar{\chi}, \bar{\psi}, \bar{\psi}')$. By (a), (b), and (c), the Cauchy stability theorem ([45], pp. 253, 255) holds, and this implies that there exists a sequence of local diffeomorphisms $F_n: M \to \bar{M}$ taking i_n to $\bar{i}$ and such that the the corresponding mapping on the metrics $F_n^*(g)$ converges to $\bar{g}$. By the argument in Ebin ([48], Proposition 6.17), using the compactness of S, there is a subsequence of F_n that converges to a local diffeomorphism $F : M \to \bar{M}$, and the domain of F is not in U. However, (M,g) and $(\bar{M},\bar{g})$ can be identified since $F_n^*(g) = \bar{g} = g$

in a neighborhood of $\overline{1}$, and the maximal development is unique. Thus (M,g) is time periodic, which is impossible.

V. Significance of the No-Return Theorem

One might think that any field theory in Euclidean space would have a non-recurrence property similar to the one demonstrated in the previous section, since a continuous field has an infinite number of degrees of freedom. From a physical point of view, this is not the case. It is not possible to make a precise measurement of the field variables at every point, and in practice a field restricted to a finite region S would be approximated by dividing up S into a finite number of subregions, and the field in each subregion replaced by its average value in that subregion. Evolution would be via the differential field equations, but one would compare the "average" values.

Comparing initial data sets via the Sobolev norm is closely analogous to comparing the average values of the field variables (and their derivatives) at one time with the average values at another time. With the Sobolev norm, one essentially takes the absolute square average value of the initial data over the entire Cauchy surface rather than dividing up the Cauchy surface into subregions; the Sobolev average is a coarser average. The arbitrary metrics e_{ab} and $\bar{g}_{ab}$ in the Sobolev norm are analogous to an arbitrary coordinate system with respect to which averages are calculated. In the classical mechanics of fields in a finite box, one can define a similar norm $\| \ \| = [\int |\Psi|^2 d\sigma]^{\frac{1}{2}}$ on the classical field Ψ. The sum $\sum_{i=1}^{n} \|\Psi^{(i)}\|$ with $\Psi^{(i)}$ the i^{th} derivative of Ψ, is essentially the same as the Sobolev norm. This sum is a map f from the space of all values of the field and its first n derivatives into $\mathbb{R}$. If the range of f is bounded - as it must be if the solutions Ψ are stable - then the evolution of Ψ must result in accumulation points in the range of f. An analogous accumulation is impossible in general relativity, as the above theorem shows. What happens in general relativity is that the range of f, so to speak, is not bounded - singularities develop in spacetime (this follows from the hypotheses of the above theorem and Theorem 2 of [45], p. 266). This sort of singular behavior is typical of many non-linear field theories [59,60], but in linear field theories such as electromagnetism, the solutions are stable, hence bounded. For example, the solutions to the Schrödinger equation in a finite box must recur[53,54] in the above average sense: for any initial value $\Psi(t_0)$ for the probability distribution and any ϵ, there exists a time t for which $\|\Psi(t) - \Psi(t_0)\| < \epsilon$.

Another argument for field theory recurrence in an average sense is Boltzmann's method of replacing the differential field equations by finite difference equations. As Boltzmann pointed out, the solutions of these difference equations would have a recurrence property.

This leads one to suspect that it might be possible to prove a Poincaré theorem for the average values of linear Hamiltonian classical fields and their first n derivatives. Unfortunately, this will not be easy to do because of the difficulty of defining the required invariant measure on the initial data space. It is known[51], for instance, that a translation invariant measure on a Hilbert space cannot give a finite value on a box in the Hilbert space. The best result to date on this problem is the definition of an infinitesimally invariant measure on the solution space to the two-dimensional Euler equation.[50]

In a finite universe without singularities there should be a finite number of physically distinguishable states (this should also be true if the matter in the universe is in the form of fields, for reasons stated at the beginning of this section). I would expect that the states of such a universe would evolve in general in a quasi-ergodic way[56,67], though it is impossible to prove this without a suitable measure. If the evolution of the physically distinguishable states is indeed quasi-ergodic, then in infinite time, every state would with high probability occur an infinite number of times.

Another way to model the evolution of such a finite state, always existing universe is by regarding the evolution as a finite discrete Markov chain [49] with stationary transition probabilities. That is, we will sample the state of the Universe at definite time intervals Δt, and we will assume that the state of the universe at time t_j is determined by the previous state at time t_{j-1} and the probability matrix of going from this previous state to any other state. If the universe has existed for an infinite time, then with probability one the states of the universe form a closed set ([49], p. 384); that is, we may consider without loss of generality the Markov chain representing the universe now to be irreducible (This is essentially Nietzsche's argument that if the universe had a final state, it would have reached it by now. It is also Birkhoff's postulate of metric transitivity [55,67]). By Theorem 4 of ref. [49] (p. 392), it follows that with probability one, all states recur in the future. Thus if Nietzsche's "chance-like" evolution is assumed to be Markovian, then his argument for recurrence is valid.

Similar arguments will imply some form of recurrence in closed universes which avoid singularities by "bouncing" at some small radius (Such behavior could of course occur only if some of the conditions in the No-return Theorem are violated[61]). One could envisage the physical constants of the universe, such as the elementary particle masses, the coupling constants, the specific entropy per baryon, and so forth, changing in some fashion at each bounce. These physical constants can be regarded as additional variables in the initial data. If the number of such constants is finite and if their range of variation is finite, then in a finite universe one would expect a finite number of physically distinguishable states. This implies accumulation points in the state space with an infinite number of bounces; with quasi-ergodic or Markovian evolution, we would have recurrence of all states with high probability. Quantum mechanical considerations do not substantially alter this conclusion, so long as one requires the number of physically distinguishable states to be finite.

It is of course possible to avoid the recurrence conclusion by assuming that the range of variation of the physical constants is not bounded, or that the universe, although closed, increases its radius at maximum expansion with each bounce. (Tolman[41], for example, argued that the monotone increase of entropy required a monotone increase of maximum radius at each bounce). However, if the range of physical constant variation is not bounded, then there would exist a sequence of cycles such that at least one physical constant would diverge in the limit. I would regard such a divergence as a singularity "at infinity". If the maximum radius increases monotonically with time, then with an infinite number of bounces in the past, the sequences of values of the maximum radius must have a limit point in the past. If its limiting value is non-zero, then one must have recurrence in the past, though this "recurrence" may take the form of a gradual cessation of change, as in the Eddington-Lemaitre universe[43]. If its limiting value is zero, then the singularity has not really been removed, it has just been placed at temporal infinity.

But the singularities in the single cycle closed universe can also be regarded as at temporal infinity. In fact, both Milne[62] and Misner[63] have contended that because physical changes occur with diverging rapidity (as measured in proper time) near a singularity, the singularity should be regarded as occurring at temporal infinity as measured in physical time. Barrow and myself[57] have pointed out that extrinsic time[64], which is a naturally defined absolute time in closed universes[65], has just this property of placing the singularities at temporal infinity. In the extrinsic time scale, a closed universe has existed and will exist forever; the question of what preceded the Big Bang does not arise.

Penrose has suggested [56] that the entropy of the gravitational field is proportional to "some suitable integrated measure of the size of the Weyl curvature," and that this curvature is zero at the initial singularity, and infinite at the final singularity. There are indications (e.g. ref.[57]) that in stable solutions with the Weyl curvature initially zero, the average Weyl curvature would increase monotonically with time to the final singularity. If this is indeed the case, the Penrose gravitational entropy could be shown to increase from the initial to the final singularity without the use of "coarse-graining", and thus could act as the ultimate source of all forms of entropy increase.

Because of Poincaré recurrence, it is impossible to define a monotonically increasing entropy (or any monotonically increasing function) for non-gravitational fields or systems of particles in terms of the phase space variables of the fields. However, since Poincaré recurrence does not hold for gravitational fields, it _is_ possible to define functions which increase monotonically with time - such functions are called Lyapounov functions[9,10] - on the phase space of the gravitational field. In fact, the extrinsic time, which is proportional to the trace of the extrinsic curvature of a leaf of a constant mean curvature Cauchy hypersurface foliation of spacetime, is just such a function since the extrinsic curvature is essentially the momentum of the gravitational field[37], and the extrinsic time increases monotonically in closed universes, ranging from $-\infty$ at the initial singularity to $+\infty$ at the final singularity[65]. In non-singular asymptotically flat space, a foliation of constant mean curvature Cauchy surfaces would have to have extrinsic time constant (= 0). This supports the conclusion of classical mechanics that isolated systems cannot define a time direction.

Niels Bohr[66] together with Prigogine and his co-workers[6,10] have suggested on the basis of Poincaré recurrence that thermodynamic concepts such as temperature and entropy are complementary to the phase space variables. In other words, they argue that a detailed microscopic description of the behavior of a physical system would preclude the definition of thermodynamic variables for that system. If the gravitational field (or more precisely, the global structure of spacetime) is taken into account, we see that it becomes possible in principle to define the thermodynamic variables and the phase space variables in a closed universe simultaneously; the notion of complementarity becomes unnecessary. This leads one to speculate about the connection between quantum mechanics and the global structure of spacetime: perhaps the impossibility of giving a detailed deterministic microscopic description of a quantum mechanical system is due to neglect of the global structure of spacetime.

Acknowledgments

I am grateful to J.D. Barrow, S.G. Brush, P.R. Chernoff, R.O. Hansen, O.E. Lanford, L. Lindblom, J.E. Marsden, and D.W. Sciama for helpful discussions. This work was supported by NSF grant PHY-77-15191.

REFERENCES

1. Kac, M. Probability and Related Topics in Physical Sciences (Interscience, London, 1959).
2. Davies, P.C.W. The Physics of Time Asymmetry (University of California Press, Berkeley, 1975).
3. Holton, G. Thematic Origins of Scientific Thought: Kepler to Einstein (Harvard University Press, Cambridge, 1973).
4. Holton, G. Science **188**, 328 (1975); Merton, R. K. Science **188**, 335 (1975).
5. Brush, S. G. The Kind of Motion We Call Heat: A History of the Kinetic Theory of Gases in the 19th Century (North Holland, Amsterdam, 1976). Volume II.
6. Prigogine, I., George, C., Henin, F. & Rosenfeld, L. Chem. Scripta **4**, 5 (1973).
7. Prigogine, I., Mayné, E., George, C., de Hann, M. Proc. Natl. Acad. Sci.USA **74**, 4152 (1977).
8. Uhlenbeck, G.E. in The Physicist's Conception of Nature (ed. Mehra, J.),p.501(Reidel, Dordrecht, 1973).
9. Prigogine, I. Science **201**, 777 (1978).
10. Misra, B. Proc. Natl. Acad. Sci. USA, **75**, 1627 (1978)
11. de Santillana, G. and von Dechend, H. Hamlet's Mill: An Essay on Myth and the Frame of Time (Gambit, Boston, 1969).
12. Baillie, J. The Belief in Progress (Oxford University Press, Oxford, 1950) p. 42.

13. Eliade, M. The Myth of the Eternal Return (Pantheon, New York, 1954).

14. Toulmin, S. and Goodfield, J. The Discovery of Time (Harper and Row, New York, 1965).

15. Sorokin, P. A. Social and Cultural Dynamics, Volume II, (American Book Company, New York, 1937).

16. Jaki, S. Science and Creation (Scottish Academic Press, Edinburg, 1974).

17. Aristotle Problemata, Book XVII, 3.

18. Plato Timaeus, 39.

19. Plato Politicus, 269 c.

20. Needham, J. Time and Eastern Man (Roy. Anthropological Institute Occasional Paper No. 21, London, 1965).

21. Thorndike, L. A History of Magic and Experimental Science During the First 13 Centuries of Our Era, Vol. 2 (Columbia University Press, New York, 1947) pp. 203, 370, 418, 589, 710, 745, 895.

22. Needham, J. Science and Civilization in China, Vol. 3 (Cambridge University Press, Cambridge, 1960).

23. Sivin, N. Earlham Review **1**, 82 (1966).

24. Tillich, P. The Protestant Era (University of Chicago Press, Chicago, 1948); see also Cullmann, O. Christ and Time, Westminister, Philadelphia, 1950).

25. Kubrin, D. J. Hist. Ideas **28**, 325 (1967).

26. Murray, J. Trans. Roy. Soc. Edinburgh **7** 411, (1815).

27. Thompson, S. P. The Life of William Thomson, Baron Kelvin of Largs (Macmillan, London, 1910), p. 111.

28. Brush, S. G. Graduate Journal 7, 477 (1967) Footnote 32 (p. 547).

29. Arrhenius, S. Worlds in the Making: The Evolution of the Universe (Harper, New York, 1908), p. 193.

30. Rankine, W. J. M. Phil. Mag. 4, 358 (1858).

31. Weinberg, S. The First Three Minutes (Basic Books, New York, 1977), p. 144.

32. Everett, A. E. Phys. Rev. D 10, 3161 (1974).

33. Poincaré, H. Acta Mathematica 13, 1 (1890); English translation in Brush, S. G. Kinetic Theory, volume 2, Irreversible Processes (Pergamon Press, Oxford, 1966).

34. Spencer, H. First Principles, 6th Edition (D. Appleton & Co., London, 1910) Chapter 23.

35. Nietzsche, F. Eternal Recurrence in Complete Works of Friedrich Nietzsche, ed. O. Levy, volume 16. (T. N. Foulis, Edinburgh, 1910).

36. Nietzsche, F. The Will to Power, volume 2, Book 4 in Complete Works of Friedrich Nietzsche, ed. O. Levy, volume 15. (T. N. Foulis, Edinburgh, 1910).

37. Misner, C. W., Thorne, K. S., and Wheeler, J. A. Gravitation (Freeman, San Francisco, 1973).

38. Boltzmann, L. Ann. Physik 59, 793 (1896); English translation in Brush, S. G. (ref. 33) p. 238.

39. Boltzmann, L. Sitzungsber. Preuss. Akad. Wiss. Berlin 1016 (1897).

40. Friedmann, A. Zeits. f. Physik 10, 377 (1922).

41. Tolman, R. C. Relativity, Thermodynamics and Cosmology, (Oxford University Press, Oxford, 1934).

42. Tolman, R. C. Phys. Rev. **38**, 1758 (1931).

43. Tipler, F. J., Clarke, C. J. S., and Ellis, G. F. R. Singularities and Horizons, in Einstein Centenary Volume, ed. A. Held (to be published, 1979).

44. Wheeler, J. A., in chapter 44 of ref. 37.

45. Hawking, S. W., and Ellis, G. F. R. The Large Scale Structure of Spacetime (Cambridge University Press, Cambridge, 1973).

46. Penrose, R. Techniques of Differential Topology in Relativity (SIAM, Philadelphia, 1972).

47. Fischer, A. E. and Marsden, J. E. The Space of Gravitational Degrees of Freedom (preprint, Berkeley, 1977).

48. Ebin, D. G. "The Manifold of Riemannian Metrics", in Proceedings of Symposia in Pure Mathematics, volume XV: Global Analysis, ed. Chern, S. S. and Smale, S. (AMS, Providence, 1970).

49. Feller, W. An Introduction to Probability Theory and Its Applications, Volume 1, 3rd edition (Wiley, New York, 1968).

50. Albeverio, S., Ribeiro de Faria, M. and Høegh-Krohn, R. Stationary Measures for the Periodic Euler Flow in Two Dimensions, (preprint, Math. Inst. (Univ. Oslo), 1978).

51. Choquet-Bruhat, Y., DeWitt-Morette, C., and Dillard-B eick, M. Analysis, Manifolds, and Physics (North-Holland, Amsterdam, 1977), p. 514.

52. Reichenbach, H. The Direction of Time (University of California Press, Berkeley, 1971).

53. Percival, I. C. J. Math. Phys. **2**, 235 (1961).

54. Bocchieri, P. and Loinger, A. Phys. Rev. **107**, 337 (1957).

55. Farquhar, I. E. Ergodic Theory in Statistical Mechanics (Interscience, London, 1964).

56. Penrose, R. in Theoretical Principles in Astrophysics and Relativity (eds. Lebovitz, N. R., Reid, W. M. and Vandervoort, P. O.)(Chicago University Press, Chicago, 1978).

57. Barrow, J. D. and Tipler, F. J. Nature **276**, 453 (1978).

58. Gibbons, G. W. and Hawking, S. W. Phys. Rev. D **15**, 2738 (1977).

59. Glassey, R.T. J. Math. Phys. **18**, 1794 (1977).

60. Levine, H. Trans. Amer. Math. Soc. **192**, 1 (1974).

61. Tipler, F.J. Phys. Rev. D **17**, 2521 (1978).

62. Milne, E.A. Modern Cosmology and the Christian Idea of God (Oxford University Press, 1952).

63. Misner, C.W. Phys. Rev. **186**, 1328 (1969).

64. York, J.W. Phys. Rev. Lett. **28**, 1082 (1972).

65. Marsden, J.E. and Tipler, F.J. preprint (1978).

66. Bohr, N. J. Chem. Soc., 349 (1932).

67. Lebowitz, J.L. and Penrose, O. Physics Today **26**, 23 (1973).

Biographical Sketch

I was born and raised in Andalusia, a small town in southern Alabama. My interest in physics dates back to my kindergarten days (circa 1952) when I became fascinated with von Braun's visions of interplanetary flight. By the time I entered M.I.T. as an 18 year old freshman in 1965, however, this interest had metamorphosed into interest in fundamental physics, with particular attention to the role of Time in scientific theories. Graduating from M.I.T. in 1969, I became a graduate student in general relativity at the University of Maryland. I obtained my Ph.D. in 1976 with a thesis on causality violation (time travel), and I have held a postdoctoral position in general relativity at the University of California, Berkeley, since 1976.

Frank J. Tipler