The Penrose-Hawking Singularity Theorems: History and Implications

John Earman

THE HISTORY OF SINGULARITY THEOREMS in Einstein's general theory of relativity (GTR) is very far from an idealized textbook presentation where an analysis of space-time singularities is followed by theorems about the existence of singularities in solutions to the Einstein field equations (EFE);[1] indeed, crucial advances in the understanding of the concept of space-time singularity were driven by a need to understand what various singularity theorems did, and did not, demonstrate. The seminal singularity theorems of Roger Penrose and Stephen Hawking relied on a new and mathematically precise definition of singularities, although this was not clear from the first publications of the results and it may not have then been entirely clear to the authors themselves. These theorems did succeed in convincing the general relativity community that singularities, in one sense of that term, are a generic feature of solutions to EFE. However, these theorems and their subsequent generalizations did not settle the debate about the correct definition of singularities; in fact, it has become increasingly clear that there is no one 'correct'

[1] Einstein's field equations (EFE), with cosmological constant term, read:

$$R_{\mu\nu} - \tfrac{1}{2} R\, g_{\mu\nu} + \Lambda\, g_{\mu\nu} = 8\pi\, T_{\mu\nu}$$

where $R_{\mu\nu}$ is the Ricci tensor, R is the Riemann curvature scalar, Λ is the cosmological constant, and $T_{\mu\nu}$ is the stress-energy tensor. An equivalent form of the field equations is

$$R_{\mu\nu} - \Lambda\, g_{\mu\nu} = 8\pi\, (T_{\mu\nu} - \tfrac{1}{2}\, g_{\mu\nu}\, T)$$

where T is the trace of $T_{\mu\nu}$. In what follows I have changed notation in the original papers to conform to the $(+ + + -)$ signature for the space-time metric. In keeping with the style of the times, I have used the component notation for tensors instead of the abstract index notation now in vogue.

H. Goenner, J. Renn, J. Ritter, T. Sauer (eds.)
The Expanding Worlds of General Relativity
(Einstein Studies, volume 7) pp. 235–267

definition and that the term 'space-time singularity' points to a sizable family of distinct though interrelated pathologies that can infect relativistic space-times.

Many general relativists view singularities as being intolerable, and those who do tend to see theorems proving the prevalence of singularities among solutions to EFE as showing that GTR contains the seeds of its own destruction. As a result there have been attempts either to modify classical GTR so as to avoid singularities or else to combine GTR with quantum mechanics in the hope that the quantum version of gravity will smooth away the singularities. A discussion of these foundations issues is beyond the scope of the present paper.[2]

The aim here is the modest one of tracing the route to the Penrose-Hawking theorems. Even so the entirety of all of the relevant literature is so large that a definitive treatment would have to be book length. The selection principle used in the present survey is to concentrate on results which contributed directly to the debate about whether or not singularities in GTR are only artifacts of the idealizations of models of cosmology and gravitational collapse. Although the present survey is far from definitive, the brevity of the treatment allows the main themes to stand out from a very cluttered background.

The singularity theorems of interest are naturally divided into four sets: a group of results by Richard Tolman, Georges Lemaître, and others from the 1930s; results by Amalkumar Raychaudhuri and Arthur Komar from the 1950s; a transitional result by Lawrence Shepley from the early 1960s; and finally the Penrose-Hawking theorems from the mid to late 1960s. The theorems from the 1930s were recognizably singularity theorems independently of the then extant controversies about how to define singularities. Nor were any of the key results from the 1930s or 1950s motivated by a desire to understand the status of singularities in the Schwarzschild solution, the De Sitter solution, or other cosmological models. It would be a mistake, however, to neglect these matters. For Raychaudhuri's first engagement with singularities in GTR stemmed from a desire to clarify the status of the $r = 2M$ Schwarzschild singularity; and, more generally, the critical reaction to the main results of Raychaudhuri and Komar reflected the unsettled state of opinion about how best to analyze singularities. Thus in what follows overviews of attempts to define and understand the nature of singularities in general-relativistic space-times will be interwoven with discussions of singularity theorems.

1. The struggle to understand space-time singularities: 1916–1939

Singularities began to demand attention soon after Einstein's general theory was codified in its final form. The Schwarzschild (1916) exterior solution was not only the first exact solution of EFE, it was also the basis of the three classical tests of the theory. In "Die Grundlagen der Physik" David Hilbert (1917) not only confronted the singularity structure of the Schwarzschild solution, but used the opportunity to give what amounts to the first general definition of singularities in GTR. In

[2] See Earman 1996 for a discussion of these issues.

coordinates introduced by Johannes Droste (1916), the line element is

$$ds^2 = \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2) - \left(1 - \frac{2M}{r}\right) dt^2. \quad (1)$$

Hilbert pronounced that this metric is singular (or "not regular") at both $r = 0$ and $r = 2M$:

> For α [$2M$] $\neq 0$, it turns out that $r = 0$ and (for positive α) also $r = \alpha$ are points at which the line element is not regular. By that I mean that a line element or a gravitational field $g_{\mu\nu}$ is *regular* at a point if it is possible to introduce by a reversible one-one transformation a coordinate system, such that in this coordinate system the corresponding functions $g'_{\mu\nu}$ are regular at that point, i.e., they are continuous and arbitrarily differentiable at the point and in a neighborhood of the point, and the determinant g' is different from 0.[3] (Hilbert 1917: 70–71)

Hilbert was, of course, correct that the apparent singularity at $r = 2M$ (the 'Schwarzschild radius') cannot be removed by a coordinate transformation that is required to be smooth and invertible not only for $r > 2M$ but also at $r = 2M$. Later critics charged that there is no need to restrict attention to coordinate transformations that are smooth at the Schwarzschild radius; on the contrary, it was said, one would expect the transformation to be singular at this radius since the Droste coordinates 'go bad' there, as evidenced by the fact that it takes an infinite amount of t-coordinate time to reach $r = 2M$ from $r > 2M$ even though only a finite amount of proper time elapses. But by itself such a remark would hardly have been found decisive; indeed, two decades later Einstein was to appeal to a similar feature of isotropic coordinates as evidence that there *is* a real singularity at the Schwarzschild radius (see below).[4]

Einstein's initial concern with the exterior Schwarzschild solution focused not on the problem of singularities but rather on the perceived anti-Machian character of the solution. As John Stachel has written, it was for Einstein a "scandal that a solution to his field equations should exist which corresponds to the presence of a single body in an otherwise 'empty universe'" (Stachel 1979: 440). The scandal seemed to worsen with the introduction of the cosmological constant term into the field equations and Willem De Sitter's (1917a, b) discovery of a solution with the line element

$$ds^2 = dr^2 + R^2 \sin^2\left(\frac{r}{R}\right)(d\chi^2 + \sin^2\psi\, d\theta^2) - \cos^2\left(\frac{r}{R}\right) dt^2, \quad (2)$$

[3] "Für $\alpha \neq 0$ erweisen sich $r = 0$ and bei positivem α auch $r = \alpha$ als solche Stellen, an denen die Maßbestimmung nicht regulär ist. Dabei nenne ich eine Maßbestimmung oder ein Gravitationsfeld $g_{\mu\nu}$ an einer Stelle *regulär*, wenn es möglich ist, durch umkehrbar eindeutige Transformation ein solches Koordinatensystem einzuführen, daß für dieses die entsprechenden Funktionen $g'_{\mu\nu}$ an jener Stelle regulär, d.h. in ihr und in ihrer Umgebung stetig und beliebig oft differezierbar sind und eine von Null verschiedene Determinante g' haben." In the second volume of *Die Relativitätstheorie* Max von Laue concurred with Hilbert (von Laue 1921: 215).

[4] An illuminating account of early attempts to understand the mysteries of the Schwarzschild solution is to be found in Eisenstaedt 1982.

where R is a positive constant (*not* the Riemann curvature scalar). The metric of this line element can be considered to be a solution to Einstein's vacuum field equations with cosmological constant $\Lambda = 3/R^2$. Here Einstein found the issues of Mach's principle and singularities joined, for he wanted to treat the violation of the former as *ersatz* on the grounds that mass concentrations were hiding in the $r = (\pi R)/2$ singularity of (2). This required an analysis of what it meant for a space-time to be singular, which was duly supplied. For Einstein, a space-time was to be counted as *non*singular if in the finite realm the covaraiant and contravariant components of the metric are continuous and differentiable and, thus, the determinant g never vanishes. A space-time point p is said to be in the finite realm if it can be joined to an arbitrary origin point p_0 by a curve of finite length. Einstein apparently had in mind using proper length along space-like or time-like curves. But if no bound is put on the acceleration of time-like curves, points 'at infinity' will be counted as lying at a finite distance. Later writers would use affine distance along a geodesic to judge what is at a finite distance. It seemed to Einstein in 1918 that the discontinuity in (2) at $r = (\pi R)/2$, which lies at a finite distance, could not be removed by any choice of coordinates and, thus, represented a real singularity.[5] In this he was mistaken, as emerged from the work of Felix Klein (1918), Kornel Lanczos (1922, 1923), and Arthur S. Eddington (1923). Referring to Einstein's idea that the De Sitter singularity represents a "'mass horizon' or ring of peripheral matter," Eddington responded that "A singular ds^2 does not necessarily indicate material particles, for we can introduce or remove such singularities by making transformations of coordinates" (Eddington 1923: 165). But then he proceeded to express his uncertainty about how to handle the general situation: "It is impossible to know whether to blame the world-structure [space-time metric] or the appropriateness of the coordinate system" (ibid.).

The following year Eddington (1924) produced a coordinate transformation that showed that the $r = 2M$ Schwarzschild singularity is a coordinate artifact, although he apparently was not aware of the significance of his result.[6] The first self-conscious demonstration of the 'fictive' nature of the $r = 2M$ singularity was due to Georges Lemaître (1932), who compared this singularity to the $r = (\pi R)/2$ singularity of the De Sitter line element (2). Although a definite advance in understanding had been achieved—not only had the nature of $r = 2M$ been clarified but the Lemaître's extension of the Schwarzschild solution also revealed the non-stationary character that extensions must possess—neither the Eddington nor the Lemaître treatment revealed the full singularity structure of the Schwarzschild solution. For, as will become clear, singularities implicate the global structure of space-time, and neither of these treatments indicate all of the relevant properties

[5] It is not clear whether Einstein, like Hilbert, meant to require that an allowed coordinate transformation must be smooth and invertible not only up to but also at the (apparent) singularity. Some evidence that he did not comes from the fact that he explicitly recognized that the $\psi = 0$ singularity in (2) is "*nur scheinbar*" because it can be removed by transforming from spherical to Cartesian coordinates; but the transformation is singular at the north pole.

[6] The transformation is $t' = t - 2M \ln(r - 2M)$. (Actually Eddington's formula was missing a factor of 2.) Note that the transformation is singular at $r = 2M$. David Finkelstein (1958) rediscovered Eddington's transformation.

of the maximal analytic extension of the exterior Schwarzschild solution.

Although Lemaître's paper appeared in two Belgian journals that were not widely read, it was known to a few key actors, including Howard P. Robertson, who had produced his own transformation for removing the $r = 2M$ singularity.[7] In 1939 Robertson was at Princeton and in contact with Einstein. In view of the fact that discussions with Robertson are explicitly acknowledged in Einstein's "On a Stationary System with Spherical Symmetry Consisting of Many Gravitating Masses" (1939), it is more than a little surprising to see how Einstein treats the $r = 2M$ singularity. Einstein chooses to write the Schwarzschild line element in isotropic coordinates:

$$ds^2 = \left(1 + \frac{\mu}{2r'}\right)(dx^2 + dy^2 + dz^2) - \left(\frac{1 - \dfrac{\mu}{2r'}}{1 + \dfrac{\mu}{2r'}}\right)^2 dt^2, \qquad (3)$$

where $r'^2 = x^2 + y^2 + z^2$, r' is related to the Droste radial coordinate r by $r = \mu + r' + \mu^2/4r'$, and μ is the mass. In this form of the line element the blow up behavior of g_{11} in (1) is removed, but as noted by Einstein, g_{44} in (3) vanishes at $r' = \mu/2$.[8]

> [This] means that a clock kept at this place would go at a zero rate. Further it is easy to show that both light rays and material particles take an infinitely long time (measured in "coordinate time") in order to reach the point $r' = \mu/2$ when originating from a point $r' > \mu/2$. In this sense the sphere $r' = \mu/2$ constitutes a place where the field is singular. (Einstein 1939: 922)

To modern eyes what this behavior indicates is not the presence of a singularity but that, as with the 'mass horizon' in the De Sitter solution (2), $r' = \mu/2$ is an event horizon.[9]

The purpose of Einstein's paper was to argue that the $r' = \mu/2$ (or $r = 2M$) Schwarzschild singularity does "not exist in physical reality." The argument given is that for the special case of a spherically symmetric cluster of particles moving in circular orbits under the influence of their mutual gravitational field, the radius of the cluster cannot be smaller than its Schwarzschild radius. It is surprising that

7 Synge (1950: 84) reports a transformation, attributed to a 1939 lecture of Robertson's, for removing the $r = 2M$ singularity.

8 I have taken the liberty of changing Einstein's r to r'.

9 Although the nature of event horizons was not clarified until much later, Lanczos (1923) was clear that $r = (\pi R)/2$ in the De Sitter solution corresponds to an event horizon and not to a singularity. J. Robert Oppenheimer and Snyder (1939) recognized that in the context of gravitational collapse the Schwarzschild radius acts as an event horizon, although they did not use this terminology; see Section 2 below. An event horizon for a system of observers is the boundary between the region of space-time from which those observers can receive causal signals and the region from which they cannot receive signals. In the maximal extension of the exterior Schwarzschild solution $r = 2M$ has an absolute status as an event horizon: it is the boundary between that portion of space-time that can be seen from future null infinity and that portion that cannot be so viewed. This is the basis of the modern definition of 'black hole.'

Einstein produced an elaborate fifteen page calculation to reach this conclusion, for it was known that in the Schwarzschild exterior field there are no circular geodesics for $r < 3M$. It is not just surprising but nearly inexplicable that Einstein thought that a static analysis, which did not allow for collapse of matter, could yield the desired impossibility result. Einstein's model, which has the particles in orbit rather than headed for a common origin, seems unconsciously chosen to yield the wanted result. In the very year that Einstein's paper appeared, Oppenheimer and Snyder (1939) showed that the gravitational collapse of matter could lead to the uncovering of the Schwarzschild radius (see Section 2).

In sum, at the end of the 1930s not only was there no agreement on how to define singularities, there was not even a consensus about the status of singularities in the key test case, the Schwarzschild solution. There were examples, such as the De Sitter solution, which showed the need to distinguish between genuine and apparent singularities; but again, there was no consensus on how this distinction was to be drawn. There was, however, at least tacit agreement on this much: if some relevant physical variable, such as energy density or a curvature scalar, becomes unbounded, and this behavior occurs at a finite distance (to use Einstein's (1918) phrase), then a genuine singularity is implicated.[10]

2. Singularity theorems of the 1930s

This Section reviews singularity theorems in the cosmological setting by Tolman and Morgan Ward (1932), Lemaître (1932), Tolman (1934a), and John Lighton Synge (1934), and results by Oppenheimer and Volkoff (1939) and Oppenheimer and Snyder (1939) for gravitational collapse.

Tolman (1930a, 1930b, 1930c) studied non-static, homogeneous, and isotropic models whose line element turned out the be equivalent to that of what are now called the Friedmann-Lemaître-Robertson-Walker (FLRW) models (see Tolman 1930d). We now know that the symmetries of the space-time metrics in question force the stress-energy tensor to have the form of a perfect fluid: $T^{\mu\nu} = (\rho + p)\, U^{\mu} U^{\nu} + p\, g^{\mu\nu}$, where ρ is the density of matter, p is the pressure, and U^{μ} is the normed four-velocity of the fluid. Tolman and Ward (1932) examined the case where the cosmological constant is set to zero and where $\rho > 0$ and $p \geq 0$. They showed that, as a consequence of EFE, if the volume of space is initially finite, there is a finite upper bound beyond which the model cannot expand; further, this bound is reached in a finite time, after which the model contracts to zero volume, also in a finite time.

This is arguably the first significant singularity theorem for GTR. Once having achieved this significant result, Tolman and Ward lost no time in trying to negate it in two ways. First, they argued that despite what the mathematical analysis says, it is plausible on physical grounds to expect that the contraction to zero volume

[10] Although I cannot cite specific passages from the literature of the 1930s to substantiate this claim, the fact that the theorems discussed in Section 2 were accepted as singularity theorems is strong indirect support for the claim.

would "be followed by renewed expansion, thus leading to a continued succession of somewhat similar expansions and contractions" (Tolman & Ward 1932: 842). This conclusion was supported by a citation to a previous paper of Tolman's which contains the stronger assertion that "it is evident physically that contraction to a zero volume could only be followed by another expansion" (Tolman 1931: 1765). The idea of an irremovable singularity—one which cannot be removed by any suitable extension of the space-time—was evidently one which Tolman did not want to contemplate.[11] Second, Tolman and Ward, citing the authority of Einstein (1931), opined that "it is possible the idealization upon which our considerations have been based should be regarded as failing in the neighborhood of zero volume" (p. 842). Specifically, they thought that the perfect fluid idealization might break down at very small volumes. Einstein, however, laid the finger of blame on symmetry assumptions. Speaking of the initial singularity in the Friedmann model, Einstein wrote: "Here one can try to get out of the difficulty by pointing out that the inhomogeneity of stellar matter makes illusory our approximate treatment" (Einstein 1931: 237).[12] Other cosmologists, such as Robertson (1932) and De Sitter (1933), joined the chorus that sang that the initial singularity in the FLRW models is an artifact of the unrealistic symmetry assumptions.

Although Tolman initially sang in harmony, he soon produced some discordant evidence. Tolman (1934a) investigated inhomogeneous dust filled models ($T^{ab} = \rho\, U^a U^b$) which exhibit spherical symmetry. Now called Tolman-Bondi models, they are more properly called Lemaître-Tolman-Bondi models since they were introduced by Lemaître (1932). Tolman (1934a) cites Lemaître, and Hermann Bondi (1947) in turn cites Tolman. The line element of this model can be written in the form

$$ds^2 = e^{\lambda}dr^2 + e^{\omega}(d\theta^2 + \sin^2\theta d\phi^2) - dt^2,$$

where λ and ω are functions of t and the radial coordinate. Applying EFE with cosmological constant yields a second order equation for ρ:

$$\frac{\partial^2 \ln \rho}{\partial t^2} = 4\pi\rho - \Lambda + \frac{1}{3}\left(\frac{\partial \ln \rho}{\partial t}\right)^2 + \frac{2}{3}\left(\frac{\dot{\omega}'}{\omega'}\right)^2, \qquad (4)$$

where the prime and dot denote respectively differentiation with respect to r and t. Consider regions where $4\pi\rho - \Lambda > 0$. If ρ is initially increasing, it follows from (4) that "reversal in the process of condensation would not occur short of arrival at a singular state involving infinite density or a breakdown of our simplified equations" (Tolman 1934a: 173). What Tolman did not say, but can be easily proved to follow from (4), is that ρ becomes infinite in a finite amount of time. Again the tendency to blame the singularity on the idealizations of the model is

---

[11] The trick is to specify what a 'suitable' extension is. If no continuity requirements are put on an extension, then any singularity is removable. For reflections on what continuity conditions are appropriate to GTR, see Earman 1995a: chap. 2.

[12] "Hier kann man der Schwierigkeit durch den Hinweis darauf zu entgehen suchen, daß die Inhomogeneität der Verteilung der Sternmaterie unsere approximative Behandlung illusorisch macht."

noteworthy. Results similar to Tolman's were obtained by Synge (1934) by means of another technique.[13]

Ancther interesting singularity result was sketched by Lemaître (1932) for a class of homogeneous but possibly non-isotropic models, which in the modern classification scheme belong to the Bianchi Type I class. The line element can be written as

$$ds^2 = \sum_{\alpha=1}^{3} g_{\alpha\alpha}(t)\, dx^\alpha dx^\alpha - dt^2. \tag{5}$$

Defining $R^3 \equiv \sqrt{-g} \equiv \sqrt{-\det(g_{ij})}$, Lemaître argued that, as a consequence of EFE without cosmological constant term, "If, at a certain moment, $\dot{R}$ is negative, it follows that R attains a value of zero and thus that the volume is annulled" (Lemaître 1932: 84).[14]

The connection between the vanishing of g and a space-time singularity is not evident at first glance;[15] but it is known from hindsight that a genuine singularity is involved. For example, in the vacuum case EFE imply that either the metric (5) is flat or else

$$g_{\alpha\alpha} = (t - t_0)^{2\sigma_\alpha} C_\alpha \qquad \text{(no summation,)}$$

where the σ_α and C_α are constants, and the σ_α must satisfy

$$\sigma_1 + \sigma_2 + \sigma_3 = 1$$
$$\sigma_1{}^2 + \sigma_2{}^2 + \sigma_3{}^2 = 1$$

(see Ryan & Shepley 1975: sect. 9.5). It follows that either the metric (5) is flat or else that there is a curvature singularity at $t = t_0$ since $R^{\mu\nu}{}_{\delta\eta}\, R^{\delta\eta}{}_{\mu\nu} \longrightarrow \infty$ as $t \longrightarrow t_0{}^+$ where $R_{\mu\nu\delta\eta}$ is the Riemann curvature tensor.

Although incomplete as a singularity theorem, Lemaître's result embodied two remarkably prescient features. First, he did not assume, as was common at the time, that matter was in the form of a dust or a perfect fluid but only that the stress-energy tensor satisfied a reasonable energy condition. Second, his argument that g goes to zero is an early form of the Raychaudhuri effect. Raychaudhuri's seminal 1955 paper will be discussed below in Section 4. It makes no reference to Lemaître. But Raychaudhuri was familiar with Lemaître's work since there is a reference to it in Raychaudhuri 1953. Thus, it is plausible that Lemaître's construction was an unconscious inspiration for Raychaudhuri's work.

Singularities reared their heads not only in cosmology but also in stellar dynamics. The most general static and spherically symmetric metric has a line element of the form

$$ds^2 = e^\lambda\, dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2 - e^\nu\, dt^2,$$

[13] For more details on Synge's technique, see Eisenstaedt 1993.

[14] "Si donc à un certain moment R' [$\dot{R}$] est négatif, il faut que R atteigne la valeur zéro et donc que le volume s'annule."

[15] As we will see in Sections 4 and 5, this problem became important in the 1950s and 1960s.

where λ and ν are functions of r alone. If matter is assumed to act as a perfect fluid, EFE without the cosmological constant term imply three ordinary differential equations:

$$\begin{cases} 8\pi p = e^{-\lambda}\left(\frac{\nu'}{r} + \frac{1}{r^2}\right) - \frac{1}{r^2} \\ 8\pi \rho = e^{-\lambda}\left(\frac{\lambda'}{r} - \frac{1}{r^2}\right) + \frac{1}{r^2} \\ p' = \frac{(p+\rho)\nu'}{2}, \end{cases} \tag{6}$$

as had been shown by Tolman (1934b). It was assumed that outside of the matter, $\rho = p = 0$, and that in this exterior region the line element takes on its Schwarzschild form

$$e^{-\lambda(r)} = 1 - \frac{2M}{r}, \qquad e^{\nu(r)} = 1 - \frac{2M}{r}.$$

Adjoining an equation of state $\rho(p)$ to (6) gives four equations for four unknowns. Oppenheimer and Volkoff (1939) used an equation of state designed to model a cold Fermi gas, which they took as a reasonable first approximation for a neutron star. They found that for masses greater than (3/4)$M_{\odot}$, static solutions for the mass distribution do not exist.

> There would then seem to be only two answers possible to the "final" behavior of very massive stars: either the equation of state we have used so far fails to describe the behavior of highly condensed matter that the conclusions reached above are qualitatively misleading, or the star will continue to contract indefinitely, never reaching equilibrium. Both alternatives deserve serious consideration. (Oppenheimer & Volkoff 1939: 380–381)

A heuristic discussion of the possible deviations from the Fermi equation of state left them confident that the first possibility was not plausible.

The need to examine non-static solutions was recognized, and in a subsequent paper Oppenheimer and Snyder (1939) provided such an analysis which bypassed the question of equation of state by studying the case of the free gravitational collapse of a dust ball ($p = 0$). In effect, the Lemaître-Tolman-Bondi solution for a spherically symmetric dust ball is being used as the interior solution joined to an exterior Schwarzschild solution.[16] It was found that after a finite proper time as measured by an observer comoving with the matter, no light signals could be sent from the star to external observers: "the cone within which a signal can escape has closed entirely" (Oppenheimer & Snyder 1939: 459). This is the first explicit and unequivocal prediction from GTR of the formation of what was to become known as a black hole. That the black hole would contain an infinite density singularity was not explicitly mentioned, but it was a clear consequence of their analysis.

[16] Oppenheimer and Synder (1939: 457) cite Tolman 1934a. But as Eisenstaedt (1993) demonstrates, it is really Lemaître who deserves the credit.

The paper ended with an acknowledgment that "actual stars would collapse more slowly than the example studied analytically because of the effect of the pressure of matter, of radiation, and of rotation" (p. 459). But they expected that the same kind of behavior would be found for "all collapsing stars which cannot end in a stable stationary state" (ibid.) A considerable effort would be needed before general agreement on this confident pronouncement could be secured. Oppenheimer and Snyder did not draw any implications from the study of gravitational collapse for the status of the $r = 2M$ Schwarzschild singularity. It was left to Raychaudhuri (1953) to make this connection (see Section 3).

At the close of the 1930s the cumulative evidence from various singularity theorems was sufficient to suggest that singularities play more than an incidental role in GTR. Yet there were few, if any, research workers who seem to have been aware of all the evidence. In the case of Lemaître's (1932) paper this is understandable since it appeared in two obscure journals. It is less easy to understand how the Oppenheimer-Volkoff and Oppenheimer-Snyder papers could be ignored since they appeared in the *Physical Review*, but ignored they were (see Section 3). And in any case the available evidence was compatible with the attitude that simplifying assumptions of symmetry and idealized forms of matter were responsible for the singularity results.

3. Further attempts to understand singularities: 1940s and early 1950s

Neither the journal literature nor textbooks from the 1940s provided much of an advance in understanding of singularities. The only reference to singularities in Peter Bergmann's *Introduction to the Theory of Relativity* (1942) is to the Schwarzschild solution, and there Bergmann endorses Einstein's (1939) claim that the $r = 2M$ singularity cannot occur in nature. However, his uncertainty about the status of the $r = 2M$ singularity is revealed by his comment on Robertson:

> Robertson has shown that, if the Schwarzschild field could be realized, a test body which falls freely towards the center would take only a finite proper time to cross the "Schwarzschild singularity," even though the coordinate time is infinite; and he has concluded that at least part of the singular character of the surface $r = 2M$ must be attributed to the coordinate system. (Bergmann 1942: 203)

A reader might well have been puzzled by this remark. How can part of the singular character of $r = 2M$ be attributed to the choice of coordinates and part not? If, as Robertson showed (see Section 2), there is a coordinate transformation which removes the $r = 2M$ singularity, is not the singularity thereby shown to be due wholly to the choice of coordinate system?[17]

Many of the features of the Schwarzschild solution were clarified in a remarkable and remarkably undercelebrated paper by John Lighton Synge (1950). From the

[17] There are two possibilities. Either Robertson did not inform Einstein and Bergmann about his transformation. Or else he did, but Bergmann thought that it did not show that the singularity could be entirely removed because the transformation is singular at $r = 2M$.

work of Lemaître and Robertson, Synge was aware that the $r = 2M$ singularity was only a coordinate artifact. He modestly described his own contribution as follows: "I have removed the Schwarzschild [$r = 2M$] singularity in a different way" (Synge 1950: 84). In fact, what Synge produced was the maximal analytic extension of the exterior Schwarzschild solution. This extension was rediscovered a decade later by C. Fronsdal (1959) and Martin Kruskal (1960), the latter of whom made an advance on Synge's analysis by presenting the metric of the maximal extension in a single global coordinate system.

Synge felt the need for a general analysis of singularities and was apologetic for not being able to supply one.

> It was hoped that at this point there might be given a brief but thorough discussion of singularities of space-time in general, and that the ideas there developed might be applied in particular to the line element [of Synge's extension]. However, the further one looks into the question of singularities, the more difficult the situation appears. ... Obviously, before we talk of singularities at all we should define them, but there are difficulties here which may not appear on the surface. ... Thus we must content ourselves for the present with definitions dependent on the coordinate system employed. (Synge 1950: 100)

Synge proceeded to define the notions of "component singularity" and "determinant singularity" in a fashion that is dependent on the choice of coordinate system and is, thus, useless from the modern point of view.

The lack of a satisfactory definition of singularity was also decried by Abraham Taub (1951), who wanted to test the validity of Mach's principle in GTR. The version of Mach's Principle at issue stated that "the nature of space-time is determined by the matter present. The latter is described either by the singularities in $g_{\mu\nu}$ ... or by the stress-energy tensor $T_{\mu\nu}$" (Taub 1951: 472). On this reading of Mach's Principle, a space-time that is empty ($T_{\mu\nu} = 0$) and is singularity-free should be flat. To conform to this principle, Einstein's field equations (without cosmological constant term) should have the property that any singularity-free solution of $R_{\mu\nu} = 0$ should be flat: $R_{\mu\nu\sigma\eta} = 0$. Taub tested this constraint for the case of solutions of $R_{\mu\nu} = 0$ admitting a three-parameter group of motions and was able to establish a restricted version of the Principle:

> **Theorem** (Taub). A spatially homogeneous space-time whose three-parameter group of motions is the group of Euclidean translations, for which $R_{\mu\nu} = 0$, and for which the $R_{\mu\nu}$ are bounded for all points with finite coordinates, is a flat space.

(This is a version of the Lemaître result mentioned in Section 2 above for Bianchi Type I models.) However, Taub also produced potential counterexamples to Mach's Principle by showing that if the components $R_{\mu\nu\sigma\eta}$ are not required to be bounded for all points with finite coordinates, then $R_{\mu\nu} = 0$ does not imply $R_{\mu\nu\sigma\eta} = 0$. But to decide whether or not the counterexamples are effective requires a decision as to whether the unboundedness of the curvature components in Taub's coordinate system corresponds to an essential singularity. But no generally accepted criterion for making the decision existed.

Amalkumar Raychaudhuri's interest in singularities was awakened by reading Bergmann's *Introduction to the Theory of Relativity* and more specifically by Bergmann's report of Einstein's (1939) attempt to show that the Schwarzschild $r = 2M$ singularity is unattainable by matter. In 1953 he showed how to join an exterior Schwarzschild field to a non-static solution of EFE representing a spherically symmetric cluster of particles moving radially towards the center of symmetry. The interior solution chosen was the Tolman form of the FLRW models. Since this form requires that the density of matter is a function of t alone, the analysis is less general than that of Oppenheimer and Snyder (1939), who allowed the density to depend on r.[18] But Raychaudhuri opined that "A spatially non-uniform distribution of particles (retaining spherical symmetry) would, however, lead to the same results so far as our investigation is concerned" (Raychaudhuri 1953: 418, n. 7). He concluded from his analysis that:

> No singularity corresponding to the [$r = 2M$] Schwarzschild singularity appears at any phase in the exterior for any arbitrary finite concentration in the cluster. The Schwarzschild singularity thus appears to be only a property of particular coordinate systems, and there seems to be no theoretical limit to the degree of concentration [of matter]. (Raychaudhuri 1953: 418).

The difference between Raychaudhuri's cluster, which can go on contracting indefinitely, and Einstein's (1939) cluster, which cannot, was explained by the fact that a "null sphere" lies beyond $r = 2M$; since Einstein's particles are supposed to move in circles, their orbits must lie outside $r = 2M$. (In fact, as noted above, they must lie outside $r = 3M$.) Lemaître (1932) had already realized that a Friedmann solution could be joined to an exterior Schwarzschild solution and, thus, that a mass could have a radius smaller than its Schwarzschild radius. In fact, this realization was what motivated him to search for a coordinate transformation that removed the $r = 2M$ singularity in the Schwarzschild-Droste coordinates.[19] An irony of the construction that led Raychaudhuri to conclude that there is a merely fictitious singularity at $r = 2M$ is that this construction also led to the prediction of a real singularity: in the collapse model "there is a singularity at a finite time, the whole region [occupied by matter] collapsing to a zero volume" (1953: 421).[20]

To summarize, a completely knowledgeable observer in the 1950s would have been able to give an account of the status of the $r = 0$ and $r = 2M$ Schwarzschild singularities. But owing to the obscurity of the journals in which the papers of

---

[18] The Oppenheimer-Snyder paper is not referenced in Bergmann's book, and consequently Raychaudhuri was unaware of it in 1953 (private communication from A. Raychaudhuri).

[19] For details, see Eisenstaedt 1993.

[20] Raychaudhuri, like most of his contemporaries, was reluctant to accept the possibility of a singularity in nature. But rather than blaming the singularity in his model on the idealizations of the analysis, he was prepared to fault GTR. "What happens after that [the infinite density singularity], our equations cannot say. It appears, indeed, that while we can trace the history of the birth of a particle, we cannot tell what happens when the particle is actually born. This perhaps can be attributed to the fact, as remarked by Einstein, that the general theory of relativity would break down under such stringent conditions" (Raychaudhuri 1953: 421). The reference here to Einstein is to *The Meaning of Relativity* (1950), where Einstein contemplated a break down of the field equations of GTR in order to avoid the big bang singularity of the FLRW models.

Lemaître (1932) and Synge (1950) appeared, few such observers existed. No progress had been made during the 1940s and early 1950s towards a general definition of singularities. Synge and Taub found this situation embarrassing, but the need for a definition was not urgent as long as it was felt that singularities in GTR arise only as artifacts of unrealistic assumptions. That impression was to be challenged by the theorems proved in the mid 1950s.

4. The singularity theorems of Raychaudhuri and Komar

Further evidence was produced in 1955 that, in the cosmological context at least, singularities in general relativistic space-times are not artifacts of symmetry assumptions. The evidence came not from one of the havens of general relativity in America or Europe, but from Calcutta. In 1953 Raychaudhuri produced an analysis of cosmological singularities that was to have a profound effect on later developments, but because it ran into difficulties with referees, the paper did not appear until 1955.[21] "Relativistic Cosmology I" contained a brief and modest abstract:

> The paper presents some general relations obtaining in relativistic cosmology. It appears from these that a simple change over to anisotropy without the introduction of spin does not solve any of the outstanding difficulties of isotropic cosmological models. (Raychaudhuri 1955: 1123)

The "outstanding difficulties" were twofold. First, the age of the universe estimated from the isotropic models and the observation of nebular distances was too short—it was not even consistent with the estimated age of the earth. Second, the models led to an "original singularity" or "creation in the finite past," which many found repugnant. The purpose of Raychaudhuri's paper was to show that these difficulties would not automatically disappear if the assumptions of homogeneity and isotropy were dropped.

Raychaudhuri's analysis assumed that cosmic matter may be treated as dust. Einstein's field equations then imply that the world lines of the dust particles are geodesics. In a coordinate system which is adapted to the dust (x^α = constant, $\alpha = 1, 2, 3$, for the dust particles) and in which coordinate time measures proper time along the geodesics, the line element is of the form

$$ds^2 = g_{\alpha\beta}\, dx^\alpha\, dx^\beta + 2g_{\alpha 4}\, dt\, dx^\alpha - dt^2 \qquad (\alpha, \beta = 1, 2, 3), \tag{7}$$

where $\partial g_{\alpha 4}/\partial t = 0$. If the $g_{\alpha 4} = 0$, the coordinate system is called 'synchronous'. In the present case where the coordinate system is adapted to the flow lines of the

21 The results were first presented in April 1953 in a letter to *Physical Review*. This letter was rejected for publication because the referee could not understand how the results were derived. A full paper was submitted to the *Astrophysical Journal*. This too was rejected, now on the grounds that it was too speculative. A modified version was sent to *Physical Review*. After hearing nothing for several months, Raychaudhuri sent queries to the editor but received no reply. After waiting for a year with no response, he sent the manuscript to *Zeitschrift für Physik*. He received a prompt rejection. Finally in February 1955 the acceptance from *Physical Review* came. The Editor, R. A. Goudsmidt, wrote: "We endeavor to choose as referees those colleagues who accept this task conscientiously. We regret in this case, there was an extensive delay." Raychaudhuri 1955 was entitled "Relativistic Cosmology I." The contemplated second part never appeared. I am most grateful to Prof. Raychaudhuri for sharing these details with me.

dust, (7) will have synchronous form if and only if the matter is nonrotating. Raychaudhuri did not assume nonrotation at this point in his analysis, but he noted that without loss of generality the $g_{\alpha 4}$ can be made to vanish along one of the flow lines of the dust matter, in which case the only non-vanishing component of the stress-energy tensor is $T_4^4 = \rho$, where ρ is the density of the dust. Einstein's field equations (with cosmological constant term) then imply that

$$R_4^4 = \Lambda - 4\pi\rho.$$

Combining this with a direct calculation of the Ricci tensor in the coordinate system of (7) yields

$$\frac{1}{G}\frac{\partial^2 G}{\partial t^2} = \frac{\Lambda - 4\pi\rho - \phi^2 + 2\omega^2}{3},$$

where $G^6 = -g$, ω is the magnitude of rotation of matter, and ϕ is a function of the metric potentials that vanishes if and only if the expansion or contraction of matter is isotropic.

In the case of non-rotating matter ($\omega = 0$), two consequences can be drawn from the previous equation, or so Raychaudhuri claimed. First, if $\Lambda = 0$, G cannot have a minimum so that "one has to start from a singularity at a finite time in the past as in isotropic models" (Raychaudhuri 1955: 1125). Second, the time scale from this singular state to the present is a maximum for isotropic models. The upshot is that dropping the assumption of isotropy does not by itself solve either of the outstanding difficulties of cosmology. However, it was also noted that the changeover to anisotropy would allow Λ to escape the lower bound on its value set by observational data in the case of isotropy, and that a higher value for Λ can give a longer time scale and even lead to an avoidance of the original singularity. But the use of this "arbitrary parameter," Raychaudhuri opined, "robs the theory of much of its appeal" (Raychaudhuri 1955: 1125).

It needs to be emphasized that Raychaudhuri's result is largely geometrical and that EFE enter at only one juncture. Consider a congruence of time-like geodesics, and let V^μ be the normed ($V^\mu V_\mu = -1$) tangent vector field of this congruence. The relevant geometrical quantities are defined as follows: the expansion $\theta \equiv \nabla_\mu V^\mu$, where ∇_μ is the derivative operator associated with the space-time metric $g_{\mu\nu}$; the shear $\sigma_{\mu\nu} \equiv \nabla_{(\mu} V_{\nu)} - (1/3)\,\theta\, h_{\mu\nu}$, where $h_{\mu\nu} \equiv g_{\mu\nu} + V_\mu V_\nu$ is the space metric of the hyperplane orthogonal to V^μ; and the rotation $\omega_{\mu\nu} \equiv \nabla_{[\mu} V_{\nu]}$. What is now called 'Raychaudhuri's equation' is a purely geometrical identity which states that

$$\dot{\theta} = -\frac{1}{3}\theta^2 - \sigma_{\mu\nu}\sigma^{\mu\nu} + \omega_{\mu\nu}\omega^{\mu\nu} - R_{\mu\nu} V^\mu V^\nu,$$

where the dot denotes differentiation with respect to proper time. Specializing to the case where there is no rotation ($\omega_{\mu\nu} = 0$), or equivalently, where the geodesic congruence is hypersurface orthogonal,[22] it follows that *if* $R_{\mu\nu}\, V^\mu V^\nu \geq 0$, then

$$\dot{\theta} + \frac{1}{3}\theta^2 \leq 0.$$

Integrating this yields the result that if the initial expansion θ_0 is negative (i.e., the geodesic congruence is initially converging), then $\theta \longrightarrow -\infty$ within a proper time $\leq 3/|\theta_0|$. Thus far Einstein's field equations have played no role. They now come on stage because they imply that $R_{\mu\nu}\, V^\mu V^\nu = 8\pi(T_{\mu\nu}\, V^\mu V^\nu + (1/2)T) + \Lambda$. If $\Lambda = 0$, then to assure that $R_{\mu\nu}\, V^\mu V^\nu \geq 0$ all that is needed in addition is that $T_{\mu\nu}\, V^\mu V^\nu + (1/2)\, T \geq 0$, which is an instance of the strong energy condition (which requires that this last inequality holds for every unit time-like V^μ). In the case of a dust model with $T_{\mu\nu} = \rho\, V_\mu V_\nu$, the strong energy condition is automatically fulfilled if ρ is non-negative, which was implicitly assumed all along. The connection with Raychaudhuri's theorem is made by noting that in a synchronous coordinate system adapted to the dust flow,

$$\theta = \frac{\partial}{\partial t} \ln(\sqrt{-g}),$$

so that $\theta \longrightarrow -\infty$ corresponds to $g \longrightarrow 0$.

But in what sense does $\theta \longrightarrow -\infty$ or, equivalently, $g \longrightarrow 0$, implicate a space-time singularity? For the moment, take $\theta \longrightarrow -\infty$ to imply the crossing of geodesics which are orthogonal to some initial space-like hypersurface. Such a crossing, however, need not indicate a genuine space-time singularity since, for example, it can happen even in Minkowski space-time with the appropriate choice of initial hypersurface.[23] In the case where matter consists of pure dust, the world lines of the dust particles are (as noted above) time-like geodesics so that crossing of these geodesics does imply an infinite density singularity. This is also seen from the conservation law $\nabla_\mu T^{\mu\nu} = 0$, which in a synchronous coordinate system adapted to the matter world lines gives $\rho\,\sqrt{-g} =$ constant, so that $g \longrightarrow 0$ implies that $\rho \longrightarrow \infty$. However, the dust idealization is unrealistic, and if, for example, pressure effects are included, matter will not follow geodesics. Furthermore, one can wonder whether even a small amount of rotation for matter will lead to an avoidance of infinite densities. Thus, in the general setting the implications of a Raychaudhuri $g \longrightarrow 0$ singularity for a space-time singularity is left unsettled.

Actually, there is a more subtle gap in the above argument that was not fully appreciated until nine years later when Shepley (1964) was able to fill the gap. One way of exposing the gap is to note that $\theta \longrightarrow -\infty$ does not necessarily mean that the geodesics cross (a specific example will be given below). The argument that, for dust matter, $\rho\,\sqrt{-g} =$ constant and, therefore, that $\rho \longrightarrow \infty$ as $g \longrightarrow 0$

[22] This equivalence follows from Frobenius' theorem. If the geodesic congruence is initially non-rotating, it will remain so.

[23] This is a point emphasized by the Russian reaction to the results of Raychaudhuri and Komar; see Section 5.

assumes the validity of the synchronous coordinate system adapted to the dust flow; but this coordinate system may break down. Of course, a new synchronous coordinate system can be resurrected in its stead if the world lines of the dust remain orthogonal to a family of space-like hypersurfaces; but this second system may break down even more quickly than the first. A third system can be erected in place of the second, but it may break down even more quickly than the second, etc. Geometrically what may go wrong is that the hypersurfaces orthogonal to the geodesic flow may change from space-like to null (again a specific example will be considered below).

Arthur Komar (1956), apparently unaware of Raychaudhuri's paper, published similar but seemingly more general results. Komar did not assume that matter is nonrotating nor that matter is in the form of dust. But he did employ a synchronous coordinate system. Geometrically such a system is derived by adapting coordinates to a congruence of time-like geodesics orthogonal to some initial space-like hypersurface; initially the rotation is zero and, one can show, will remain zero. This congruence may or may not represent the flow of matter. Using a synchronous coordinate system, Komar defined a symmetric tensor field (now called the extrinsic curvature) $\chi_{\alpha\beta} \equiv \partial g_{\alpha\beta}/\partial t$ and showed that if $\Lambda = 0$ and $T_{44} + (1/2)\, T \geq 0$, χ^{α}_{α} will diverge at a finite time unless it is initially zero. (The relation to Raychaudhuri's result is seen from the fact that in a synchronous coordinate system, $\chi^{\alpha}_{\alpha} = \partial \ln \sqrt{-g}/\partial t$.) What did Komar take the significance of this formal result to be? The title of the paper, "Necessity of Singularities in the Solutions of the Field Equations of General Relativity," was explained in the introductory section.

> The question naturally arises whether such singular points [as the initial singularity in the Friedmann model] are a consequence of the particular symmetry proposed in Friedmann's model, or whether perhaps for more general distributions of matter one need not expect instants of creation or annihilation of the universe. The purpose of this paper is to show that singularities in the solution of the field equations of general relativity are to be expected under very general hypotheses ... (Komar 1956: 544)

In the penultimate section of the paper this confident pronouncement was taken back with the acknowledgment that the connection between the divergence of χ^{α}_{α} and "instants of creation or annihilation" is not pellucid.

> We should note that one cannot easily determine whether the singularity is in the coordinate system or whether the space itself is singular. Taub[24] has pointed out that there is as yet no well-defined way of determining what constitutes an essential singularity within the general theory of relativity. (Komar 1956: 546)

And yet in the conclusion of the paper the initial pronouncement is partially restated.[25] The condition $\chi^{\alpha}_{\alpha} = 0$, together with the other assumptions of the analysis, imply that space-time is flat. Thus, if space-time is not flat,

[24] The reference here is to Taub 1951; see Section 3 above.

[25] It seems plausible to conjecture that the reference to Taub was added in the revised version of the paper prepared in reaction to the referee's report.

> we must be prepared either: (A) to allow for singularities (as in the Schwarzschild or Friedmann solutions); (B) to permit the possibility of a cosmological term or a negative pressure term ... ; or (C) to consider spaces which do not have the simplifying property of containing a set of geodesically parallel space-like hypersurfaces for all times. (Komar 1956: 546)

Komar was correct in dropping the worry about whether the singularity is only in the coordinate system. Although his argument is framed in coordinate terms, the point is purely geometrical. If non-flat solutions to Einstein's field equations are considered, Λ is set to 0, and $T_{\mu\nu} V^\mu V^\nu + (1/2) T \geq 0$ for all time-like V^μ, then the space-time cannot be covered by a 'geodesically parallel' family of space-like hypersurfaces. For if there were such a family, that would mean that there is an everywhere defined congruence of non-rotating time-like geodesics. One could then apply the Raychaudhuri-Komar effect to the unit four-velocity field V^μ of the geodesic congruence to conclude that the geodesics cross at some finite time, yielding a contradiction, at least if it is assumed that the geodesics of the congruence can be prolonged indefinitely. To repeat, this result does not assume that matter is in the form of dust nor that matter is non-rotating—the V^μ need not be the four-velocity of matter.[26] But for this very reason the generality of Komar's result is achieved at the expense of severing the connection with singularities in the sense of an infinite matter density since now $\theta = \nabla_\mu V^\mu \longrightarrow -\infty$ need not imply a matter density singularity.

To probe further the significance of Komar's result, suppose that in order to escape the contradiction Komar derived, one drops either the assumption that space-time is covered by a one-parameter family of geodesically parallel space-like hypersurfaces or that the geodesics of normal congruence can be extended indefinitely far (as measured in affine distance). What does this portend for singularities in the sense then in play? Taub-NUT space-time[27] is a vacuum solution to Einstein's field equations with $\Lambda = 0$. It trivially satisfies the strong energy condition, and it is non-flat. It is covered by a one-parameter family of hypersurfaces on which the metric is homogeneous. The family members covering the initial or Taub portion of the space-time are space-like. But the member at the boundary of the Taub and NUT regions is null. And yet there is no singularity in the sense with which Raychaudhuri, Komar, and their interlocutors were concerned—there is no infinite density of matter (the Taub-NUT universe being a vacuum solution), nor is there any curvature blow up. Furthermore, this example shows why taking $\theta \longrightarrow -\infty$ to mean that the geodesics cross is only a loose way of speaking. For the time-like geodesics orthogonal to an initial space-like surface of homogeneity in the Taub portion of the space-time, θ does become infinitely negative as the Taub-NUT boundary is approached; but these geodesics do not cross but rather asymptote to a null hypersurface. As a result the geodesics cannot be extended to

26 In a note responding to Komar's paper, Raychaudhuri (1957) claimed that Komar's result was "explicitly given by the present author [i.e., Raychaudhuri 1955]". This is true if Raychaudhuri's construction is interpreted in the modern way as applying to any time-like geodesic congruence and not just to the world lines of dust matter.

27 See Misner 1963 and Misner & Taub 1968 for a description of this space-time.

indefinitely large values of affine parameter. In this sense Taub-NUT space-time *is* singular. But the adoption of this conception of singularities required a new way of thinking about singularities, a way that was and still is controversial (see Sections 6 and 9).

These points were certainly not clear to the researchers at the time. Nevertheless, the Raychaudhuri and Komar results were attacked, but from another direction.

5. The Russian reaction to the theorems of Raychaudhuri and Komar

The specter of singularities rampant among general relativistic cosmological models had been raised by the results of Raychaudhuri and Komar. The Russian school of Lev Landau, Yevgeniy Lifshitz, and coworkers sought to slay it. Before soft pedaling the implications of the Raychaudhuri-Komar results, the Russian school laid a priority claim. Lifshitz and Khalatnikov (1960b) claimed that Landau had "long ago" proved that in a synchronous coordinate system the determinant g of the metric vanishes within a finite time.[28] They went on to note that the vanishing of g need not indicate a singularity in the space-time itself but only a breakdown in the coordinate system due to the crossing of the geodesics orthogonal to some initial space-like hypersurface. For the Russian school, a "true physical singularity in the metric" is "one which belongs to the space-time itself and is not connected with the character of the reference system." By this they meant that "Such a singularity is characterized by scalar quantities, such as density of matter and the invariants of the curvature tensor, becoming infinite" (Lifshitz & Khalatnikov 1963: 190–191).[29] To settle the question of whether singularities are a general feature of solutions to Einstein's field equations, the Russian school proposed to implement the following program. First, write down the form of the most general solution of the field equations in the neighborhood of a singularity. Second, count the number of arbitrary functions of coordinates in such a solution. Third, count the number of arbitrary functions needed to fix an initial distribution of matter and the state of the free gravitational field. Fourth and finally, compare the two counts and conclude that singularities are not a general feature of general relativistic space-times just in case the latter count is larger, meaning that the subset of singular solutions is of 'measure zero' in the full set of solutions. By the Russian reckoning, the number of arbitrary functions of coordinates in a singular solution is always one less than for a general solution. They thus felt justified in stating:

---

[28] They refer to Komar 1956 but not to Raychaudhuri 1955, perhaps because the proof given by the Russian school is closer in style to Komar's. Lifshitz & Khalatnikov 1963 does refer to "an analogous result of Raychaudhuri."

[29] The Russians were not alone in this characterization of true or physical singularities. Thus, John Graves and Dieter Brill spoke of the need to distinguish between "true geometric singularities at which invariants of the Riemann curvature tensor become singular, and 'pseudo-singularities,' which are due to an unfortunate choice of coordinate system" (1960: 1507). As noted above, more careful writers might have added that the blow up behavior must occur 'at a finite distance,' the point being that if the blow up only happens as spatial or temporal infinity is approached, no singularity in the space-time itself is indicated. However, Penrose's concept of naked singularities covers behavior which happens, so to speak, only at infinity; see Section 9 below.

> All the foregoing leads to the fundamental conclusion that the presence of a time singularity is not an essential property of the cosmological model of the general theory of relativity, and the general case of an arbitrary distribution of matter and gravitational field does not lead to the appearance of a singularity. (Lifshitz, Sudakov, & Khalatnikov 1961: 1301)

This claim was repeated in a communication to *Physical Review Letters* (Khalatnikov, Lifshitz, & Sudakov 1961), in a review article by Lifshitz and Khalatnikov (1963), and in the revised second edition of *Classical Theory of Fields* by Landau and Lifshitz (1962).

The Russians eventually recanted, but not until after the singularity theorems of Penrose (1965), Hawking (1965, 1966a–d) and Robert Geroch (1966). A silent recantation took place with the 1967 Russian edition of *Classical Theory of Fields* which omitted § 110 ("The absence of singularities in the general cosmological solution"). A more explicit recantation come in 1970 in the form of a communication to *Physical Review Letters*. Khalatnikov and Lifshitz frankly admitted the limitations and pitfalls of their method:

> Since there exists no systematic method for examining the singularities of the solutions of Einstein's equations, our search for increasingly more general solutions of this kind proceeded essentially by trial and error. A negative result from such a procedure could of course never be completely conclusive by itself ... (Khalatnikov & Lifshitz 1970: 78)

As will become clear later in our story, the Russian recantation has a curious flavor because the Russian school was concerned with one sense of singularity while the theorems of Penrose et al. were concerned with a different sense.

6. Analysis of singularities in the early to mid-1960s

The difficulties in interpreting the results of Raychaudhuri and Komar pointed to the need for a better understanding of the elusive concept of space-time singularity. A plea for more clarity on these matters was made by Charles Misner in 1963. Misner's analysis was to have a crucial influence on the development of singularity theorems, though the influence lay as much in what was ignored in Misner's analysis as what was taken from it.

> The first step is to find some clearly stated problems, and the clue to clarity is to refuse even to speak of a singularity but instead to phrase everything in terms of the properties of differentiable metric fields on manifolds. If one is given a manifold and on it a metric which does not at all points satisfy the necessary differentiability requirements, one simply throws away all the points of singularity. The starting point for any further discussion is then the largest submanifold on which the metric is differentiable. ... The first problem then is to select a criteria which will identify in an intuitively acceptable way a "non-singular space." ... The problem ... is to recognize the holes left in the space where singular (or even regular) points have been omitted. (Misner 1963: 924)

For a Riemannian space $(M, k_{\mu\nu})$ with a positive definite metric, the recognition of "holes" can be achieved by investigating the metric topology. If $d(p, q)$, $p, q \in$

M, is defined as the greatest lower bound on the $k_{\mu\nu}$ length of paths joining p and q, it is easily seen that $(M, d(,))$ obeys the axioms for a metric space. The deletion of points leaving behind holes can then be detected by the incompleteness of the distance function $d(,)$ in the sense that not every Cauchy sequence of points converges.[30] For a relativistic space-time $(M, g_{\mu\nu})$ this criterion of incompleteness is unworkable since $g_{\mu\nu}$ does not define a distance function. However, in the Riemann case, Cauchy completeness is equivalent to geodesic completeness, the latter meaning that any geodesic can be extended to an arbitrarily great value of an affine parameter. This suggests using geodesic completeness as the defining characteristic of a non-singular space-time. Misner partially endorsed this suggestion by taking geodesic completeness as a *sufficient* condition for non-singularity of a space-time. His grounds were that a geodesically complete space-time is maximal, i.e., cannot be isometrically imbedded as a proper subset of a larger space-time. But this only shows that the space-time in question cannot have been obtained from a larger space-time by deleting *regular* points and leaves open the possibility that *singular* points were deleted. In fact, some later analyses counted some geodesically complete space-times as singular.[31] On the other hand, Misner rejected geodesic completeness as a *necessary* condition for non-singularity of a space-time. No compact manifold (without boundary) can be imbedded as a proper subset of another (Hausdorff) manifold of the same dimension. Thus, for either a compact Riemann space or a compact relativistic space-time, there are no holes that arise from deleting regular or singular points. For the Riemann case all is well since if M is compact, $(M, k_{\mu\nu})$ is geodesically complete. But Misner showed that all is not well for the relativistic space-time case by producing an example of a space-time $(M, g_{\mu\nu})$ that is geodesically incomplete despite the fact that M is compact.[32] More generally, if the incomplete geodesics are contained in a compact set, then the space-time is not counted as singular according to Misner's point of view; in particular, Taub-NUT space-time is seen as non-singular (see Misner & Taub 1968). Misner also acknowledged the generally accepted sufficient condition for a singular space-time; namely, a curvature scalar becomes unbounded along an incomplete geodesic. But he also warned that "It is not to be presumed that all spaces which should be called 'essentially singular' are identified by this criteria" (Misner 1963: 926).

The upshot of Misner's discussion was partly encouraging and partly discouraging. The encouraging part was that there is a clear project: for $(M, g_{\mu\nu})$ with $g_{\mu\nu}$ defined differentiable (to whatever degree you like) at all points of M, define what it means for $(M, g_{\mu\nu})$ to be non-singular. The discouraging part was that

[30] p_i, $i = 1, 2, 3, \ldots$ is a Cauchy sequence just in case, for any $\epsilon > 0$, there is an N such that for any $m, n > N$, $d(p_m, p_n) < \epsilon$.

[31] In the b-boundary approach of Bernd Schmidt, a space-time is counted as singular if it contains inextendible curves of finite generalized affine length (see Hawking & Ellis 1973: 283–284). This can happen even if the space-time is geodesically complete.

[32] Examples of this sort had been discussed by Lawrence Marcus (1962), from whom Misner said that he had "borrowed heavily" (Misner 1963: 924, n. 4). Another example was produced by Robert Hermann (1964).

although there are a number of ideas waiting to be used, there was no obvious way to combine these ideas into a precise, relatively simple, and intuitively appealing definition.

György Szekeres (1960), like Synge (1950) and Taub (1951) before him, also decried the lack of an adequate definition of a singularity of a Lorentzian manifold. He implicitly embraced part of the sentiments Misner was to express three years later; namely, the starting point for analysis is a differentiable manifold M and a Lorentz metric $g_{\mu\nu}$ defined and C^{∞} at every point of M. But Szekeres not only wanted to say what it is for $(M, g_{\mu\nu})$ to be singular (or non-singular); for singular space-times he also wanted to be able to speak of singularities, and took the first steps towards a local characterization of these objects. Using an equivalence relation on geodesics, he defined a boundary point of the space-time as an equivalence class of incomplete geodesics. Such a boundary point was called a singularity just in case it remains a boundary point in every extension of the space-time. Singularities were then classified as 'ordinary' (or 'non-ordinary') according as some derivative (respectively, no derivative) of the metric in a normal coordinate system along a geodesic fails to approach a limit as the singularity is approached. Szekeres' paper seem to have passed virtually unnoticed by those engaged in debate of the early to mid-1960s over the existence and nature of singularities in GTR.[33]

In sum, in the mid-1960s, as in the 1950s and the beginning of the 1960s, there existed no adequate and generally accepted analysis either of the notion of a space-time singularity or a singular space-time. One might think that this lacuna would make it difficult to prove general theorems about the existence of singularities in solutions to Einstein's field equations—for how can a mathematical theorem be established for a mathematically ill defined object? Such a question presupposes a naive view of how science actually works. In the event, uncertainties about the definition of singularities allowed room for maneuvering. After the dust had settled, one could work backwards from the theorems to a definition of singularities. This is not necessarily a self-aggrandizing procedure, at least not if the theorems are beautiful enough and powerful enough. They were.

7. The singularity theorem of Shepley

While the Russians were trying to exorcise the specter of singularities raised by the Raychaudhuri-Komar results, Lawrence Shepley, a student of John Wheeler and Charles Misner, was working to make the specter more threatening. Like Raychaudhuri, Shepley (1964) focused on dust models but of a sort that allowed the dust to be rotating. The models were assumed to be spatially homogeneous in the sense that the space-times admit the three-dimensional Lie group SO(3, $\mathbb{R}$) as a symmetry.[34]

[33] Geroch 1968a is the only relevant reference from the period that I have found to cite Szekeres 1960.

[34] SO(3, $\mathbb{R}$) is the group of unit determinant, 3×3 orthogonal, real matrices. The $k = +1$ (spatially closed) FLRW models belong to this class. But the class is much broader in that it includes anisotropic models.

Any such space-time is covered by a one-parameter family $H(t)$ of surfaces (which are topologically S^3) of homogeneity. It was further assumed that initially the $H(t)$ are space-like. If they remained forever space-like then a contradiction would result by the Raychaudhuri-Komar construction since the $H(t)$ are geodesically parallel. It would seem that the only way out is to conclude that for some value of t, $H(t)$ turns from being space-like to null, as in the Taub-NUT example. But Shepley established that such a change in character of the $H(t)$ is not possible for SO(3, $\mathbb{R}$) dust models. Thus, for a special class of models, Shepley showed that one of the gaps in the Raychaudhuri-Komar analysis could be filled.

At first glance, however, Shepley's result is puzzling. With the gap filled, why doesn't a genuine contradiction result, showing that such models are impossible? Shepley took the escape hatch to be the existence of singularities. He stated his result as

> **Theorem** (Shepley). All dust filled cosmological models of general relativity which have the symmetry SO(3, $\mathbb{R}$) and which have at least one space-like invariant three-dimensional hypersurface have a point singularity.

But what sense of singularity is indicated, and exactly how does the existence of such a singularity resolve the contradiction?

The answer lies in the fact that the contradiction results from assuming that the geodesics normal to the initial space-like hypersurface of homogeneity can be prolonged to an arbitrarily large value of an affine parameter. The escape hatch lies in the conclusion that the space-time is singular in the sense of having incomplete time-like geodesics. Here we have the first air-tight geodesic incompleteness singularity theorem, although some hindsight is needed to produce this reading of Shepley's paper. Shepley defines a "point of singularity" as a point "which can be reached by a geodesic of finite total length from other points of the space-time manifold, where the metric is degenerate or otherwise irregular (for example, a point where a curvature scalar is infinite)" (1964: 1403). The definition suggests, but does not say, that geodesic incompleteness is *the* identifying characteristic feature of a singular space-time. Having helped to prepare Misner's 1963 paper, Shepley was aware of both the attractions and the pitfalls of such a suggestion.[35] And possibly as a result, Shepley's definition is a superposition of the old ideas of curvature blow up and geodesic incompleteness.

As noted by Shepley himself, the connection of his singularity result to a singularity in the infinite density sense is far from evident. The conservation law $\nabla_\mu(\rho V^\mu V^\nu) = 0$ yields $\rho V^4\sqrt{-g}$ = constant in a synchronous coordinate system. In the Raychaudhuri case of non-rotating matter, the synchronous coordinate system can be chosen to be comoving with the matter and $V^4 = -1$, so that, if the synchronous coordinate system remains valid, $g \longrightarrow 0$ implies that ρ diverges. But in Shepley's more general case of rotating dust, it could conceivably happen that V^4 becomes infinite while ρ stays finite.

[35] Misner 1963: 924, n. 4 thanks Shepley for preparing the review of the literature on singularities.

8. The singularity theorems of Penrose, Hawking, and Geroch

The major turning point in the study of singularities was undoubtedly Penrose's (1965) article "Gravitational Collapse and Space-Time Singularities." The importance of this article is belied by its brevity—it occupied less than three pages of *Physical Review Letters*. Although Penrose's argumentation was somewhat obscure (as will be discussed shortly), it quickly became clear to the experts that singularities of gravitational collapse could no longer be dismissed as artifacts of the symmetries or the special conditions on matter assumed in the Oppenheimer-Volkoff-Snyder analysis (see Section 2). The article also set off a flurry of activity that, within a few short years, resulted in the generally accepted opinion that, if Einstein's field equations are correct, singularities are to be expected in generic circumstances in both gravitational collapse and cosmology.

Because of the importance of Penrose's theorem, it is worth going through the assumptions in some detail. It was assumed, first, that the space-time $(M, g_{\mu\nu})$ is temporally orientable so that the null half-cones can be continuously divided into 'past' and 'future.' This is not a restrictive assumption since if it fails for $(M, g_{\mu\nu})$ it can be secured by passing to a covering space-time. A second, and very strong, assumption is that $(M, g_{\mu\nu})$ possesses a Cauchy surface (a space-like hypersurface which is intersected exactly once by every time-like curve without endpoint) that is non-compact. Spatially closed universes are thus excluded from the purview of the theorem. And the Cauchy property precludes any hint of acausal structure for the space-time; it is, for example, even stronger than the property of stable causality which says (intuitively) that the null cones can be widened out without closed causal loops resulting.[36] Third, it is assumed that $R_{\mu\nu} K^{\mu} K^{\nu} \geq 0$ for any null vector K^{μ}. If Einstein's field equations hold, with or without cosmological constant, this condition will be fulfilled as long as matter obeys the condition $T_{\mu\nu} K^{\mu} K^{\nu} \geq 0$ for any null vector K^{ν}, which by continuity is a consequence of the weak energy condition: $T_{\mu\nu} V^{\mu} V^{\nu} \geq 0$ for all time-like V^{μ}. (Penrose's paper stipulated that $(R_{\mu\nu} - (1/2)\, R\, g_{\mu\nu} + \Lambda\, g_{\mu\nu}) V^{\mu} V^{\nu} \geq 0$ for all time-like V^{μ}, which is stronger than the theorem required.) Fourth, there must be a trapped surface $\mathcal{T}$, i.e., a space-like two-surface such that the outgoing and ingoing null geodesics that intersect it orthogonally are both converging. In the Oppenheimer-Volkoff-Snyder model of gravitational collapse, such a surface will form when the body contracts within its Schwarzschild radius—the two-sphere $r =$ constant, $t =$ constant for $r < 2M$ being an example. But—and this is the crucial point—the concept of a trapped surface does not presuppose the spherical symmetry of the Oppenheimer-Volkoff-Snyder analysis. Finally, according to Penrose, it needs to be assumed that the space-time is future null geodesically complete. "The existence of a singularity can never be inferred ... without an assumption such as completeness of the manifold under consideration" (Penrose 1965: 58).

36 For a precise definition, see Hawking & Ellis 1973: 198. The existence of a Cauchy surface is a necessary condition for the global version of determinism, according to which initial data on a time slice uniquely fix the entire future (and past).

When Penrose summarized his results as showing that "*deviations from spherical symmetry cannot prevent space-time singularities from arising*" (Penrose 1965: 58) the reader in 1965 would naturally have taken this to mean that given the above assumptions, it follows that gravitational collapse eventuates in a physical singularity in the sense understood in the debate over the prevalence of physical vs. coordinate singularities—i.e., infinite density or an infinite curvature scalar. The caption of Penrose's figure 1, showing what is presumably a density or curvature singularity in spherical gravitational collapse, also encouraged this interpretation: "The diagram essentially serves for the discussion of the asymmetrical case" (Penrose 1965: 58, fig. 1). In fact, however, nothing about a density or curvature singularity is part of or is proved by the theorem. Null geodesic completeness is an assumption of the proof only in the reductio sense. In the now accepted reconstruction of the theorem (see Hawking & Ellis 1973: 263–264 and Wald 1984: 239–240) the *reductio* assumption is used to prove that the boundary of $I^+(\mathcal{T})$, the chronological future of the trapped surface,[37] is compact. Then the compactness of $I^+(\mathcal{T})$ is shown to contradict the existence of a non-compact Cauchy surface. Thus, given the other assumptions, the space-time is null geodesically incomplete. We have a singularity theorem if and only if null geodesic incompleteness is taken as a sufficient condition of a singular space-time.

Six months after Penrose's seminal paper appeared in print, Hawking and George Ellis (1965) submitted a note to *Physics Letters* containing a new singularity theorem that generalized Shepley's (1964) result for homogeneous cosmologies. Shepley had assumed dust matter; Hawking and Ellis treated the more general case of a perfect fluid. Shepley had assumed spatial homogeneity in the form of partition of space-time by a one-parameter family of hypersurfaces of homogeneity; Hawking and Ellis assumed only that there is at least one space-like hypersurface on which a three-parameter group of motions acts transitively. Following Penrose's lead, Hawking and Ellis also assumed that the models in question were null and time-like geodesically complete: "Any models in which this were not the case would not seem to be reasonable models of the universe" (Hawking & Ellis 1965: 246). They then proceeded to demonstrate the existence of a "physical singularity" in the form of an infinite matter density. Of course, if the matter density becomes unbounded "at a finite distance," the space-time structure breaks down and geodesic completeness results. Unlike Penrose's (1965) result, however, the Hawking and Ellis theorem is not a pure *reductio* proof of geodesic incompleteness since it supplies the reason for the incompleteness.

The unclarity over what was being demonstrated in the singularity theorems persisted for at least another few months. On 16 August 1965 *Physical Review Letters* received a communication from Hawking entitled "Occurrence of Singularities in Open Universes." Hawking noted that in open ($k = 0$ or $k = -1$) FLRW models there are trapped surfaces. He argued that universes that are similar on a large scale to FLRW universes but are not homogeneous or isotropic locally would

[37] For a space-time $(M, g_{\mu\nu})$ and a set $S \subset M$, $I^+(S)$ is defined as the set of all $p \in M$ such that there is a future directed time-like curve from S to p.

still possess a trapped surface. Penrose's (1965) theorem could then be applied to deduce the existence of a singularity. This theorem was characterized by Hawking as follows: "Penrose has shown that either a physical singularity must occur or space-time is incomplete if there is a closed trapped surface" (Hawking 1965: 689).

The 22 August 1966 issue of *Physical Review Letters* published communications from Hawking (1966a) and Geroch (1966) containing new singularity theorems. Both authors were now quite specific that geodesic incompleteness is to be taken as a defining characteristic of a singular space-time—time-like incompleteness for Hawking, time-like or null incompleteness for Geroch. When giving their definitions of a singular space-time both authors referred to Misner 1963, which is somewhat ironic since one of the purposes of Misner's paper was to argue that geodesic incompleteness is *not* a sufficient condition for labeling a space-time as singular (see Section 6). And while taking geodesic incompleteness as the definition of a singular space-time served to clarify the otherwise murky logic of the previous singularity theorems—Shepley 1964, Penrose 1965, Hawking & Ellis 1965, and Hawking 1965—and to make the new theorems mathematically precise, it also undercut the claim that what was being demonstrated was the existence of "physical singularities" in the sense which was then in play in the literature and which had been the core of the debate about the prevalence of singularities in solutions to Einstein's field equations. Nevertheless, the shift of focus to geodesic incompleteness was entirely justified as a piece of opportunism. The techniques of Penrose, Hawking, and Geroch could be used to prove rigorous and powerful results about geodesic incompleteness. *Pace* Misner, even if geodesic incompleteness does not always signal singularities in the originally intended sense, it surely is a pathology worth noting; and one could suppose that in typical circumstances this pathology would be a symptom of density or curvature singularities. But these were matters that could be sorted out later. Opportunism demanded that the theorems be proved; their exact physical significance would become apparent in the fullness of time.

The Penrose-Hawking singularity results were extended and codified in a series of three articles by Hawking, published in the *Proceedings of the Royal Society* (Hawking 1966b, 1966c, 1967); in Hawking's (1966d) Adams Prize Essay "Singularities and the Geometry of Space-Time"; in Penrose's (1966) Adams Prize Essay "An Analysis of the Structure of Space-Time";[38] in a joint article by Hawking and Penrose (1970); in Penrose's (1972) monograph *Techniques of Differential Topology in General Relativity*; and in Hawking and Ellis' (1973) *The Large Scale Structure of Space-time.* The Raychaudhuri-Komar effect plays an important role in some of the theorems, but now this effect is explicitly recognized to function as part of a *reductio* demonstration of geodesic incompleteness rather than as part of a demonstration of a 'physical' singularity.

---

[38] Portions of this essay were published in Penrose (1968).

9. What is a space-time singularity?

The Penrose-Hawking theorems focused attention on geodesic incompleteness as the mark of a singular space-time. In part, there was a good reason for this choice: it provides a mathematically precise criterion that corresponds to intuitive judgments in a number of paradigm cases. The choice was also partly a matter of expediency: the proof techniques developed by Penrose and Hawking lent themselves to this definition. But, powerful and elegant as they are, the Penrose-Hawking theorems did not settle the debate about the correct definition of space-time singularity. A strong indication of the situation shortly after several of the key theorems had appeared in print was the publication of Geroch's (1968b) "What Is a Singularity in General Relativity?" The body of the paper is in the form of a Galilean dialogue. Although such a format is unusual for *Annals of Physics*, it was nicely tailored to revealing the uncertainties and ambiguities which existed at the time. What Geroch's article and subsequent analysis revealed was a situation of daunting complexity.[39]

Begin, as Misner (1963) recommended, with a relativistic space-time $(M, g_{\mu\nu})$, where M is a differentiable manifold and $g_{\mu\nu}$ is a Lorentz signature metric which is defined and differentiable on all of M. There are then two tasks. The first is to find a criterion that will detect when $(M, g_{\mu\nu})$ is singular (despite the fact that $g_{\mu\nu}$ is everywhere well-defined and smooth and, in that sense, nonsingular). If one wants to speak not only of a singular space-time but also of singularities, then the second task is to provide a means that would justify talking about these things as localizable objects. This would involve constructing a set of idealized points—to represent the singularities—and (at least) a topology for the manifold M + idealized points. Unfortunately, extant procedures yield counterintuitive results, e.g. the singular points may not be Hausdorff separated from the interior points of M; and there is reason to believe that such results will be common to all procedures which conform to some seemingly natural requirements (see Geroch, Liang, & Wald 1982). It remains to be seen whether or not 'object talk' about singularities can be given an expression that is at once mathematically precise, intuitively appealing, and useful in classifying singular space-times.

Returning to the prior task of demarcating singular space-times, four or more families of ideas compete for attention. The first starts from the intuitive idea that motivated most of the pre-Penrose-Hawking singularity theorems; namely, a singular space-time is one in which a relevant physical quantity blows up. But the technical elaborations of this idea have gone far beyond anything considered by the pioneers of singularity theorems. For instance, a space-time may be considered to be singular even if all curvature scalar polynomials remain bounded if some component of the Riemann curvature tensor, as measured in an orthonormal frame parallelly propagated along a geodesic, becomes unbounded at a finite affine distance.

[39] Technical elaborations of the concepts mentioned in this Section can be found in a number of sources: Hawking & Ellis 1973; Ellis & Schmidt 1977; Wald 1984; Scott & Szekeres 1994. For an overview of the literature, see Earman 1995a: chap. 2.

A second family of ideas identifies a singular space-time as one which exhibits some form of incompleteness. Geodesic incompleteness—the criterion used in the Penrose-Hawking theorems—belongs to this family, but there are many other members. Geroch (1968b) produced an example of a space-time which is time-like, space-like, and null geodesically complete but which contains inextendible time-like curves of bounded acceleration and finite proper length. Such a curve might, for example, correspond to the world line of a rocket ship whose motor uses only a finite amount of fuel. The pilot of such a ship might, not unnaturally, complain that he inhabits a singular space-time. An even more demanding notion of completeness can be formulated using the concept of generalized affine length which is available for all differentiable curves, not just geodesics or curves of bounded acceleration.

There is obviously a solid connection in one direction between these first two families of ideas: the unboundedness of, say, a curvature scalar at a finite distance, as measured by affine length along a geodesic, entails geodesic incompleteness. But the implication in the other direction can fail. An interesting example of the failure is to be found in the dispute between Ludwik Silberstein and Einstein,[40] although the principals were apparently unaware of the fact. Silberstein (1936) claimed to have produced a stationary and axi-symmetric solution to the source-free EFE, with two singularities lying on the otherwise singularity-free axis of symmetry. If correct, this claim would have been an embarrassment for Einstein's program of deriving the geodesic postulate from the field equations by treating a test particle as a singularity of the field, for due to the stationary character of the solution the two "mass centers" do not move towards one another, in contradiction to "man's most ancient primitive experience" (Silberstein 1936: 270). Einstein's response was that Silberstein's solution (which was in fact the Curzon (1924a, 1924b) bipolar solution) is singular along the axis joining the "mass centers" (Einstein & Rosen 1936). He was correct in that the metric cannot be smoothly extended to cover the axis, and consequently, the solution is geodesically complete. However, all the components of the Riemann curvature tensor remain well-behaved as the axis is approached.[41]

This gap between the first two families also serves to undercut the sincerity of the Russian school's recantation of the claim that singularities are absent from a general solution to EFE. In reaction to the singularity theorems of Penrose et al., Khalatnikov and Lifshitz wrote:

> The situation has changed since the discovery of Penrose (and later by Hawking and Geroch), of new theorems which reveal a connection between the existence of a singularity (of unknown type) and some very general properties of the equations, which bear no relation to the choice of reference system. (Khalatnikov & Lifshitz 1970: 78)

---

40 See Havas 1993 for a detailed account of this controversy.

41 Neither Einstein nor Silberstein nor Nathan Rosen realized that the singularities of the Curzon-Silberstein solution are not of the simple pole type but have a complicated topological structure. The bizarre global structure of the Curzon monopole solution was untangled only recently by Susan Scott and Peter Szekeres (1986).

The paper ended with a charming flourish: "The new developments finally clarify the problem of singularities in general solutions and remove all previous contradictions" (Khalatnikov & Lifshitz 1970: 78). Marx could have told them otherwise. The dialectic of singularities was only getting in full swing. Khalatnikov and Lifshitz were explicit that they were concerned with singularities in the sense of "infinite density of matter or (in empty space) infinite curvature invariants" (Khalatnikov & Lifshitz 1970: 76). The theorems of Penrose et al. did not deal with singularities in this sense but rather in the sense of geodesic incompleteness. In the year before the recantation in *Physical Review Letters*, Belinskiy and Khalatnikov once again advanced considerations that argued "in favor of the absence of a physical singularity in the general cosmological solution of Einstein's equations" (Belinskiy & Khalatnikov 1969: 911). They mentioned Penrose's theorem that "there exists (under very natural assumptions) a singularity." But, they added, this singularity is one "whose character, however, no one has succeeded in establishing and which, apparently, is so weak that it does not appear in the invariants of the curvature tensor" (Belinskiy & Khalatnikov 1969: 911). Under special conditions geodesic incompleteness can be shown to entail the blow up of a component of the Riemann curvature tensor in a parallelly propagated frame (see Clarke 1975). But the general conditions under which one can move from singularities in the sense of incompleteness to curvature singularities remains obscure.

A third family of ideas is rooted in Misner's (1963) notion that a nonsingular space-time is one without any holes or missing points. Because of the lack of the appropriate technical elaboration of this notion,[42] it is not possible at the present time to make any general statements about the relations to the first two families, except to say that they will surely turn out to be complex. Misner's (1963) examples of cases of incomplete geodesics contained in compact sets show that a space-time can be singular in the incompleteness sense even though there are no missing points.

A fourth family of ideas is centered on Penrose's notion of a naked singularity, which in turn can be defined as a violation of 'cosmic censorship.' Cosmic censorship comes in many varieties, the strongest of which requires that a space-time contain a Cauchy surface. On this version of cosmic censorship anti-De Sitter space-time is counted as nakedly singular even though it is geodesically complete and displays no curvature blow up.[43] Contrary to Einstein's (1918) idea of a singularity, what happens at infinity as well as what happens at a finite distance determines whether a naked singularity is present. In the other direction, the big bang singularity of the FLRW models is not counted as nakedly singular even though it is highly visible and involves curvature blow up and geodesic incompleteness. The hypothesis that naked singularities do not develop from regular initial data in generic solutions of EFE was put forward over a quarter of a century ago by Penrose (1969). Despite all of the effort devoted to it, the cosmic censorship

[42] The work of Scott and Szekeres (1994) can be seen as lending itself to this task.

[43] See Hawking & Ellis 1973: 131–134 for a description of this space-time.

hypothesis remains controversial.[44]

The failure of strong cosmic censorship is a failure of causality: unless a space-time possesses a Cauchy surface, Laplacian determinism in its global form cannot hold. There are many other forms of causal pathologies which can be exhibited by solutions to EFE, up to and including the existence of closed time-like curves. These causal pathologies may, or may not, be associated with curvature blow up or geodesic incompleteness. Attempts have been made to prove that, consistent with EFE and energy conditions, closed time loops cannot be produced by a finite device. Such theorems (e.g., Hawking 1992), while formally correct, are less than convincing as proof of the impossibly of operating a 'time machine' (see Earman 1995b) in GTR.

In the light of the developments sketched above, it seems pointless to try to produce a simple formula that will count as *the* correct definition of a singular space-time. When we try to explore our naive conception of singularities in the setting of relativistic space-times, we encounter a wide array of phenomena. We are still far from an understanding of the interrelations of these phenomena and the roles they play in GTR. In that sense, the Penrose-Hawking theorems are a starting point rather than the endpoint for the study of singularities in GTR. The conclusion of Geroch's Ph.D. dissertation is as appropriate today as it was thirty years ago:

> What a strange little object is the singularity with its strange properties and nonexistent definition. Yet the singularity promises to remain one of the most intriguing and disturbing aspects of gravitation theory for a long time to come. Here is a problem with which we must some day come to grips—at least if we are ever to understand this phenomenon called gravitation. (Geroch 1967: 145)

Acknowledgements

I am grateful to Jean Eisenstaedt, Al Janis, and Andrzej Krasiński for helpful comments on an earlier draft of this paper.

References

Belinskiy, V. A. & Khalatnikov, I. M. (1969). "On the Nature of the Singularities in the General Solution of the Gravitational Equations." *Journal of Experimental and Theoretical Physics (USSR)* 56: 1701–1712; *Soviet Physics JETP* 29: 911–917.

Bergmann, Peter G. (1942). *Introduction to the Theory of Relativity*. New York: Prentice-Hall.

Bondi, Hermann. (1947). "Spherically Symmetric Models in General Relativity." *Monthly Notices of the Royal Astronomical Society* 107: 410–425.

Clarke, C. J. S. (1975). "Singularities in Globally Hyperbolic Space-Time." *Communications in Mathematical Physics* 41: 65–78.

[44] See Earman 1995a: chap. 3 for an overview on the pros and cons of cosmic censorship.

CURZON, H. E. J. (1924a). "Bipolar Solutions of Einstein's Gravitation Equations." *Proceedings of the London Mathematical Society* 23: xxix.

——— (1924b). "Cylindrical Solutions of Einstein's Gravitation Equations." *Proceedings of the London Mathematical Society* 23: 477–480.

DE SITTER, Willem. (1917a). "On the Curvature of Space." *Koninklijke Akademie van Wetenschappen te Amsterdam. Proceedings of the Section of Sciences* 20: 229–242.

——— (1917b). "On Einstein's Theory of Gravitation and its Astronomical Consequences. Third Paper." *Monthly Notices of the Royal Astronomical Society* 78: 3–28.

——— (1933). "On the Expanding Universe and the Time Scale." *Monthly Notices of the Royal Astronomical Society* 93: 628–634.

DROSTE, Johannes. (1916). "The Field of a Single Center in Einstein's Theory of Gravitation, and the Motion of a Particle in that Field." *Koninklijke Akademie van Wetenschappen te Amsterdam. Proceedings of the Section of Sciences* 19: 197–215.

EARMAN, John. (1995a). *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*. New York: Oxford University Press.

——— (1995b). "Outlawing Time Machines: Chronology Protection Theorems." *Erkenntnis* 42: 125–139.

——— (1996). "Tolerance of Spacetime Singularities." *Foundations of Physics* 26: 263–640.

EDDINGTON, Arthur Stanley. (1923). *The Mathematical Theory of Relativity*. Cambridge: Cambridge University Press.

——— (1924). "A Comparison of Whitehead's and Einstein's Formulæ." *Nature* 113: 192.

EINSTEIN, Albert. (1918). "Kritisches zu einer von Hrn. De Sitter gegebenen Lösung der Gravitationsgleichungen." *Sitzungsberichte der Königlichen Preussischen Akademie der Wissenschaften* (Berlin): 270–272.

——— (1931). "Zum kosmologischen Problem der allgemeinen Relativitätstheorie." *Sitzungsberichte der Preussischen Akademie der Wissenschaften* (Berlin): 235–237.

——— (1939). "On a Stationary System with Spherical Symmetry Consisting of Many Gravitating Masses." *Annals of Mathematics* 40: 922–936.

——— (1950). *The Meaning of Relativity*. 4th ed. London: Meuthen.

EINSTEIN, Albert & ROSEN, Nathan. (1936). "Two-Body Problem in General Relativity Theory." *Physical Review* 49: 404–405.

EISENSTAEDT, Jean. (1982). "Histoire et singularités de la solution de Schwarzschild (1915–1923)." *Archive for History of Exact Sciences* 27: 157–198.

——— (1993). "Lemaître and the Schwarzschild Solution." In *The Attraction of Gravitation: New Studies in the History of General Relativity* (Einstein Studies, vol. 5). J. Earman, M. Janssen and J. Norton, eds. 353–389. Boston: Birkhäuser.

ELLIS, George F. R. & SCHMIDT, Bernd G. (1977). "Singular Space-Times." *General Relativity and Gravitation* 8: 915–953.

FINKELSTEIN, David. (1958). "Past-Future Asymmetry of the Gravitational Field of a Point Particle." *Physical Review* 110: 965–967.

FRONSDAL, C. (1959). "Completion and Embedding of the Schwarzschild Metric." *Physical Review* 116: 778–781.

GEROCH, Robert P. (1966). "Singularities in Closed Universes." *Physical Review Letters* 17:445–447.

——— (1967). "Singularities in the Space-Time of General Relativity: Their Definition, Existence, and Local Characterization." Ph.D. Thesis, Princeton University.

——— (1968a). "Local Characterization of Singularities in General Relativity." *Journal of Mathematical Physics* 9: 450–465.

——— (1968b). "What is a Singularity in General Relativity?" *Annals of Physics* 48: 526–540.

GEROCH, Robert P., LIANG, C. & WALD, Robert M. (1982). "Singular Boundaries of Space-Times." *Journal of Mathematical Physics* 23: 432–435.

GRAVES, John C. & BRILL, Dieter R. (1960). "Oscillatory Character of Reissner-Nordström Metric for an Ideal Charged Wormhole." *Physical Review* 120: 1507–1513.

GRISHCHUK, L. P. (1966). "Some Remarks on the Singularities in the Cosmological Solutions of the Gravitational Equations." *Journal of Experimental and Theoretical Physics (USSR)* 51: 475–481; *Soviet Physics JETP* 24 (1967): 320–324.

HAVAS, Peter. (1993). "The General Relativistic Two-Body Problem and the Einstein-Silberstein Controversy." In *The Attraction of Gravitation: New Studies in the History of General Relativity.* (Einstein Studies, vol. 5). J. Earman, M. Janssen and J. D. Norton, eds. 88–125. Boston: Birkhäuser.

HAWKING, Stephen W. (1965). "Occurrence of Singularities in Open Universes." *Physical Review Letters* 15: 689–690.

——— (1966a). "Singularities in the Universe." *Physical Review Letters* 17: 444–445.

——— (1966b). "The Occurrence of Singularities in Cosmology." *Proceedings of the Royal Society of London A* 294: 511–521.

——— (1966c). "The Occurrence of Singularities in Cosmology II." *Proceedings of the Royal Society of London A* 295: 490–493.

——— (1966d). "Singularities and the Geometry of Space-Time." *Adams Prize Essay*, mimeo..

——— (1967). "The Occurrence of Singularities in Cosmology III." *Proceedings of the Royal Society of London A* 300: 187–201.

——— (1992). "Chronology Protection Conjecture." *Physical Review D* 46: 603–611.

HAWKING, Stephen W. & ELLIS, George F. R. (1965). "Singularities in Homogeneous World Models." *Physics Letters* 17: 246–247.

——— (1973). *The Large Scale Structure of Space-Time.* Cambridge: Cambridge University Press.

HAWKING, Stephen W. & PENROSE, Roger. (1970). "The Singularities of Gravitational Collapse and Cosmology." *Proceedings of the Royal Society of London A* 314: 529–548.

HERMANN, Robert. (1964). "An Incomplete Compact Homogeneous Lorentz Metric." *Journal of Mathematics and Mechanics* 13: 497–501.

HILBERT, David. (1917). "Die Grundlagen der Physik: zweite Mitteilung." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen*: 53–76.

KHALATNIKOV, I. M. & LIFSHITZ, Yevgeniy M. (1970). "General Cosmological Solution of the Gravitational Equations with a Singularity in Time." *Physical Review Letters* 24: 76–79.

KHALATNIKOV, I. M., LIFSHITZ, Yevgeniy M. & SUDAKOV, V. V. (1961). "Singularities of the Cosmological Solutions of Gravitational Equations." *Physical Review Letters* 6: 311–313.

KLEIN, Felix. (1918). "Über die Integralform der Erhaltungssätze und die Theorie der räumlich-geschlossenen Welt." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen*: 394–423.

KOMAR, Arthur. (1956). "Necessity of Singularities in the Solution of the Field Equations of General Relativity." *Physical Review* 104: 544–546.

KRUSKAL, Martin D. (1960). "Maximal Extension of Schwarzschild Metric." *Physical Review* 119: 1743–1745.

LANCZOS, Kornel. (1922). "Bemerkung zur De Sitterschen Welt." *Physikalische Zeitschrift* 23: 539–543.

——— (1923). "Über die Rotverschiebung in der Sitterschen Welt." *Zeitschrift für Physik* 17: 168–189.

LANDAU, Lev D. & LIFSHITZ, Yevgeniy M. (1965). *Classical Theory of Fields.* Revised 2nd ed. Reading, MA: Addison-Wesley.

VON LAUE, Max. (1921). *Die Relativitätstheorie.* Vol. 2. Braunschweig: F. Vieweg.

LEMAÎTRE, Georges. (1932). "L'Univers en expansion." *Publication du Laboratoire d'Astronomie et de Géodésie de l'Université de Louvain* 9: 171–205. Also in *Société Scientifique de Bruxelles. Annales A* 53 (1933): 51–85.

LIFSHITZ, Yevgeniy M. & KHALATNIKOV, I. M. (1960a). "On the Singularities of Cosmological Solutions of the Gravitational Equations I." *Journal of Experimental and Theoretical Physics (USSR)* 39: 149–157; *Soviet Physics JETP* 12 (1961): 108–113.

——— (1960b). "On the Singularities of Cosmological Solutions of the Gravitational Equations." *Journal of Experimental and Theoretical Physics (USSR)* 39: 800–808; *Soviet Physics JETP* 12 (1961): 558–563.

——— (1963). "Investigations in Relativistic Cosmology." *Advances in Physics* 12: 185–249.

LIFSHITZ, Yevgeniy M., SUDAKOV, V. V. & KHALATNIKOV, I. M. (1961). "Singularities of Cosmological Solutions of Gravitational Equations. III." *Journal of Experimental and Theoretical Physics* (USSR) 40: 1847–1855; *Soviet Physics JETP* 13 (1961): 1298–1303.

MARCUS, Lawrence. (1962). Mimeographed lecture notes, American Mathematical Society Summer Institute, Santa Barbara (CA).

MISNER, Charles W. (1963). "The Flatter regions of Newman, Unti, and Tamburino's Generalized Schwarzschild Space." *Journal of Mathematical Physics* 4: 924–937.

MISNER, Charles W. & TAUB, Abraham H. (1968). "A Singularity-Free Empty Universe." *Journal of Experimental and Theoretical Physics (USSR)* 55: 233–255; *Soviet Physics JETP* 28 (1969): 122–133.

OPPENHEIMER, J. Robert & SNYDER, H. (1939). "On Continued Gravitational Contraction." *Physical Review* 56: 455–459.

OPPENHEIMER, J. Robert & VOLKOFF, G. M. (1939). "On Massive Neutron Cores." *Physical Review* 54: 374–381.

PENROSE, Roger. (1965). "Gravitational Collapse and Space-Time Singularities." *Physical Review Letters* 14: 57–59.

——— (1966). "An Analysis of the Structure of Space-Time." *Adams Prize Essay.*

——— (1968). "Structure of Space-Time." In *Battelle Rencontres. 1967 Lectures in Mathematics and Physics.* C. M. DeWitt and J. A. Wheeler, eds. 121–235. New York: W. A. Benjamin.

——— (1969). "Gravitational Collapse: The Role of General Relativity." *Revisita del Nuovo Cimento,* Serie I, 1, Numero Speciale: 252–276.

——— (1972). *Techniques of Differential Topology in General Relativity.* Philadelphia: SIAM.

RAYCHAUDHURI, Amalkumar. (1953). "Arbitrary Concentrations of Matter and the Schwarzschild Singularity." *Physical Review* 89: 417–421.

——— (1955). "Relativistic Cosmology I." *Physical Review* 98: 1123–1126.

——— (1957). "Singular State in Relativistic Cosmology." *Physical Review* 106: 172–173.

ROBERTSON, Howard P. (1932). "The Expanding Universe." *Science* 76: 221–226.

RYAN, Michael P. & SHEPLEY, Lawrence C. (1975). *Homogeneous Relativistic Cosmologies.* Princeton (NJ): Princeton University Press.

SCHWARZSCHILD, Karl. (1916). "Über das Gravitationfeld eines Massenpunktes nach der Einsteinschen Theorie." *Sitzungsberichte der Königlichen Preussischen Akademie der Wissenschaften* (Berlin): 189–196.

Scott, Susan & Szekeres, Peter. (1986). "The Curzon Singularity II. Global Picture." *General Relativity and Gravitation* **18**: 571–583.

——— (1994). "The Abstract Boundary—a New Approach to Singularities on Manifolds." *Journal of Geometry and Physics* **13**: 223–253.

Shepley, Lawrence C. (1964). "Singularities in Spatially Homogeneous, Dust-Filled Universes." *Proceedings of the National Academy of Sciences* **52**: 1403–1409.

Silberstein, Ludwik. (1936). "Two-Centers Solution of the Gravitational Equations, and the Need for a Reformed Theory of Matter." *Physical Review* **49**: 268–270.

Stachel, John. (1979). "The Genesis of General Relativity." In *Einstein Symposion, Berlin* (Lecture Notes in Physics, vol. 100). H. Nelkowski, A. Hermann, H. Posner, R. Schrader and R. Seiler, eds. 428–442. Berlin: Springer-Verlag.

Synge, John Lighton (1934). "On the Expansion or Contraction of a Symmetrical Cloud Under the Influence of Gravity." *Proceedings of the National Academy of Sciences* **20**: 635–640.

——— (1950). "The Gravitational Field of a Particle." *Proceedings of the Royal Irish Academy* **53**: 83–114.

Szekeres, György. (1960). "On the Singularities of a Riemannian Manifold." *Publicationes Mathematicæ* (Debrecen) **7**: 285–301.

Taub, Abraham H. (1951). "Empty Space-Times Admitting a Three-Parameter Group of Motions." *Annals of Mathematics* **53**: 472–490.

Tolman, Richard C. (1930a). "The Effect of the Annihilation of Matter on the Wave-Length of Light from the Nebulæ." *Proceedings of the National Academy of Sciences* **16**: 320–337.

——— (1930b). "More Complete Discussion of the Time-Dependence of the Non-Static Line Element for the Universe." *Proceedings of the National Academy of Sciences* **16**: 409–420.

——— (1930c). "On the Estimation of Distances in a Curved Universe with Non-Static Line Element." *Proceedings of the National Academy of Sciences* **16**: 511–520.

——— (1930d). "Discussion of Various Treatments Which Have Been Given to the Non-Static Line Element for the Universe." *Proceedings of the National Academy of Sciences* **16**: 582–594.

——— (1931). "On the Theoretical Requirements for a Periodic Behavior of the Universe." *Physical Review* **38**: 1758–1771.

——— (1934a). "The Effect of Inhomogeneity on Cosmological Models." *Proceedings of the National Academy of Sciences* **20**: 169–176.

——— (1934b). *Relativity, Thermodynamics and Cosmology*. Oxford: Oxford University Press.

Tolman, Richard C. & Ward, Morgan. (1932). "On the Behavior of Non-Static Models of the Universe When the Cosmological Constant Term Is Omitted." *Physical Review* **39**: 835–843.

Wald, Robert M. (1984). *General Relativity*. Chicago: University of Chicago Press.