The historical foundation of Einstein's general theory of relativity : a case study in scientific heuristics /

Author:
Norton, John Daniel

Publication Date:
1981

DOI:
https://doi.org/10.26190/unsworks/7335

THE HISTORICAL FOUNDATION OF EINSTEIN'S GENERAL THEORY OF RELATIVITY

A Case Study in Scientific Heuristics

by

John Daniel Norton

A dissertation to be submitted to the University of New South Wales for consideration for the degree of Doctor of Philosophy.

School of History and
Philosophy of Science

1981

ABSTRACT

In this dissertation I review and analyse the origins, development and prospects of the heuristics fundamental to Einstein's discovery of his general theory of relativity. They include his early conviction that an acceptable gravitation theory could not be contained within his special theory of relativity, but was to be sought in an extended theory of space and time; his postulation of the equivalence of acceleration and gravitation as a bridge over which he could pass from inertial to gravitational fields; and his epistemological critique of space, which demanded its final elimination from physical theory as a causally active agent.

While each of these heuristics contributed decisively to the final form of the theory, I argue that Einstein's adoption of them was either ill-supported or even directly contradicted by the theorems and concepts of the final theory. However, we can still regard the emergence of the final theory as the rational outcome of a single process. Einstein was driven forward by the demands of these productive but volatile heuristics, but his progress was checked and directed by a constraint envelope which arose from the requirement that his final theory contain a relativistically acceptable gravitation theory. It is my view that this alone guaranteed that he could ultimately converge only to his final theory or one very much like it. Einstein refused to violate the requirements of this constraint envelope, even when he thought it required him to sacrifice the universality of his principle of general covariance. The power of such a constraint envelope to generate a theory of the nature of the final general theory of relativity is illustrated with the historical case study of the development of Nordström's scalar theory of gravitation in the 1912-1914 period.

"In the light of knowledge attained, the happy achievement seems almost a matter of course, and any intelligent student can grasp it without too much trouble. But the years of anxious searching in the dark, with their intense longing, their alternations of confidence and exhaustion and the final emergence into the light - only those who have experienced it can understand that."

Albert Einstein
"Notes on the Origin of the General Theory of Relativity"

PREFACE

This dissertation contains many English translations of the original German of Einstein and his contemporaries. Where recognised English translations are available, I have made use of them and this has been noted in the appropriate footnote. Otherwise, the translations are my own and the German source passage is given in full in the appropriate footnote.

I am indebted to my supervisor Dr. John Saunders for his valuable guidance and patient support and to Dr. Guy Freeland for shouldering the adminstrative burden of supervision; to the School of History and Philosophy of Science and its staff and students for providing me with an intellectual home and Dr. David Oldroyd in particular for always having time; to Mr. John Shepanski for guiding me through some of the intricacies of the general theory of relativity; to Professor Simon Prokhovnik for his encouragement and most stimulating discussions; to Professor John Stachel for his most useful correspondence and for sending me copies of his valuable work; to Mr. Jozsef Illy for his kindness in sending me copies of his detailed and lengthy manuscripts; to the inter-library loan staff at the University of New South Wales library for tirelessly tracking down the most impossible of requests; to Professor George Szekeres for his encouragement, discussion and help with translations on reading the final draft; and to Susan Kellar for her painstaking typing of the final manuscript.

Finally I must thank my mother and father-in-law for their patient advice on innumerable fine points of translation and my wife Eve for her assistance in the task of proof reading. This, of course, cannot even begin to acknowledge the many ways in which my family and, in particular, my wife have encouraged and supported me throughout this project.

TABLE OF CONTENTS

page

SYMBOLS AND NOTATION

Throughout this dissertation, I have endeavoured to use a single, standard and transparent system of notation. Because of the need to present and discuss Einstein's work using the original formalism, it has been necessary to deviate from this standard notation in many places. They are all clearly indicated so that no confusion should arise. However, in the list which follows, I have indicated a number of symbols which are not standard by noting the chapter in which they appear in parentheses.

In general, four dimensional scalars, vectors and tensors are represented by "A", "A_μ", "$A_{\mu\nu}$" etc. Latin indices vary over 1,2 and 3 and indicate spatial components. Greek indices usually vary over 0,1,2 and 3, with 0 indicating the time component. Where Einstein's notation is followed, Greek indices vary over 1,2,3 and 4, with 4 indicating the time component. The distinction between raised (contravariant) and lowered (covariant) indices is preserved even in special relativistic equations, unless otherwise indicated. Spacetime of special relativity is assumed to have a Lorentzian metric $ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu = c^2dt^2 - dx^2 - dy^2 - dz^2$ unless otherwise indicated. The Einstein summation convention is used throughout, again unless otherwise indicated. Partial differentiation is indicated by a comma: "$A_{\mu,\nu}$" and covariant differen tiation by a semi-colon: "$A_{\mu;\nu}$".

Three dimensional vectors are written as "$\underline{a}$", with absolute value "a" and components in the x,y and z directions of a Cartesian coordinate system "a_x", "a_y" and "a_z" respectively. Unit vectors in these three directions are "$\underline{i}$", "$\underline{j}$" and "$\underline{k}$" respectively.

a	acceleration
a_{μ}^{ν}	transformation matrix
A	constant
B	constant
β	$1/\sqrt{1-v^2/c^2}$
c	speed of light
c_L	speed of light in (Lorentzian) gravitation free space
c_o	constant
$dV, d\tau$	three dimensional volume element ($d\tau$: ch.7)
d^4x	four dimensional volume element
D	constant
$\bar{D}$	trace of stress-energy tensor of non-electromagnetic and non-gravitational matter. (ch.8)
δ	constant
δ	variation operator
δ_{μ}^{ν}	=1 if $\mu=\nu$; =0 if $\mu\neq\nu$
Δ	increment
Δ	Laplacian operator
$\Delta\theta$	angle of deflection
$\overline{\text{div}}$	divergence operator
E	constant
E	energy
E_o	total energy of system (ch.8)
$\mathcal{E}$	charge
$\eta_{\mu\nu}$	Lorentzian values of metric tensor
f, f (φ)	function, function of φ
F_{μ}	four-force
$\mathcal{F}$	force density (ch.7)
$\underline{f}$	three dimensional force

g	constant acceleration
g	gravitation factor, converts inertial mass to gravitational mass (ch.8)
g	determinant of $g_{\mu\nu}$
$g_{\mu\nu}$	metric tensor
$g^{*}_{\mu\nu}$	deviations of metric tensor from Lorentzian values
G	gravitational constant
G	trace of gravitational field stress-energy tensor (ch.8)
$G_{\mu}, G_{\mu\nu}, G_{\mu\nu\rho}$	field tensors for scalar, vector and second rank tensor field theories
$G_{\mu\nu}$	Einstein tensor, $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$
γ	constant acceleration
$\gamma_{\mu\nu}$	contravariant metric tensor (ch.7)
$\Gamma_{\mu\nu}$	gravitational field tensor (ch.7)
grad	gradient operator
i	$\sqrt{-1}$
$\underline{i}$	unit vector in x-direction
I	integral to be varied
$\underline{j}$	unit vector in y-direction
k	constant in field equations
k	$= ct$, path parameter
$\underline{k}$	unit vector in z-direction
(K_x, K_y, K_z)	three dimensional force vector (ch.3)
$\varkappa$	constant in field equations
l	length of a rod
L, L'	x component of magnetic field strength vector (ch.3)
L	trace of electromagnetic field stress-energy tensor (ch.8)
L	$= \frac{ds}{dt}$, Lagrangian.

m	rest mass
m_o	mass constant
M	source mass, sun's mass
M, M'	y component of magnetic field strength vector (ch.3)
M_μ^ν	non-gravitational stress-energy tensor, excluding stress terms
μ	rest mass (ch.3)
N, N'	z component of magnetic field strength vector (ch.3)
ν	frequency of light
ν	rest mass density (ch.8)
ω	gravitational contraction/dilation factor
p^μ	$= m \frac{dx^\mu}{d\tau}$, four-momentum of a body with rest mass m
π	pressure
$\Phi, \varphi, \varphi_\mu, \varphi_{\mu\nu}$	scalar, vector and second rank tensor gravitational potentials
φ_a	gravitational potential in region remote from that under examination
φ'	scaled gravitational potential
q	speed (ch.3)
r	radial distance coordinate in polar coordinate system, $r^2=x^2+y^2+z^2$
$\underline{R}$	gravitational field strength vector (ch.2)
R_μ^g	proper density of four-force (ch.8)
$R^\alpha{}_{\beta\gamma\delta}$	Riemann curvature tensor
$R_{\mu\nu}$	$= R^\alpha{}_{\mu\nu\alpha}$, Ricci tensor
R	$= R^\alpha_\alpha$, curvature scalar
$(\mathcal{R}_{x_a}, \mathcal{R}_{y_a}, \mathcal{R}_{z_a})$	non-gravitational force (ch.7)
ρ	rest mass density
rot	curl operator
s	interval in metric
S, Σ	coordinate system
S_μ^ν	stress terms in non-gravitational stress-energy tensor

σ	proper time as read by a clock (ch.4,5)
σ	rest mass density (ch.7)
Σ	summation
t,t'	time coordinate
t	clock period (ch.8)
t_{μ}^{ν}, $\theta_{\mu\nu}$	gravitational field stress-energy tensor ($\theta_{\mu\nu}$:Ch.7)
T^{μ}	$= \rho U^{\mu}$, vector source term
T_{μ}^{ν}, $\Theta_{\mu\nu}$	non-gravitational stress-energy tensor ($\Theta_{\mu\nu}$:ch.7)
T	$= T_{\alpha}^{\alpha}$, trace of non-gravitational stress-energy tensor
τ ,T	proper time
τ	proper time at origin of coordinates (ch.4,5)
τ	time coordinate (ch.4,5)
u	ict
U^{μ}	$= \frac{dx^{\mu}}{d\tau}$, four-velocity
$\underline{v}$	velocity
v	speed
(v_x, v_y, v_z)	x,y and z components of velocity $\underline{v}$
x,x'	Cartesian space coordinate
x^{μ}, x'^{μ}	general spacetime coordinate, t,x,y,z as $\mu = 0,1,2,3$ etc.
X,X'	x component of electric field strength vector (ch.3)
X_g, X'_g	x component of gravitational field strength vector (ch.3)
ξ	space coordinate, corresponds to x coordinate
y,y'	Cartesian space coordinate
Y,Y'	y component of electric field strength vector (ch.3)
Y_g, Y'_g	y component of gravitational field strength vector (ch.3)
η	space coordinate, corresponds to y coordinate
z, z'	Cartesian space coordinate

Z, Z' z component of electric field strength vector (ch.3)

Z_g, Z'_g z component of gravitational field strength vector (ch.3)

ζ space coordinate, corresponds to z coordinate

$\underline{\nabla}$ $= \underline{i}\frac{\partial}{\partial x} + \underline{j}\frac{\partial}{\partial y} + \underline{k}\frac{\partial}{\partial z}$

∇^2, Δ $= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$, Laplacian operator

$\Box^2, \Box$ $= \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2$, d'Alembertian operator

$\left\{ {\alpha \atop \mu\nu} \right\}$ $= \frac{1}{2} g^{\alpha\beta}(g_{\mu\beta,\nu} + g_{\beta\nu,\mu} - g_{\mu\nu,\beta})$, Christoffel symbols of the second kind

Primes denote that a quantity is measured in the primed coordinate system.

Overhead dots indicate time derivatives: $\dot{x} = \frac{dx}{dt}$ etc.

Subscript "o" denotes a proper or self measured quantity.

Subscript "r" denotes a quantity evaluated in the rest frame of reference.

GLOSSARY OF TERMS

The terms defined below may be used frequently in accounts of relativity theory. Their meaning may vary from account to account. In order to avoid the possibility of confusion, I define what the meaning of these terms is taken to be for the purposes of this dissertation.

The Equivalence of Acceleration and Gravitation asserts that all physical effects caused by an acceleration of the frame of reference are indistinguishable from the physical effects of a gravitational field of the appropriate strength and direction. The hypothesis is most often limited to uniform acceleration and homogeneous gravitational fields, where this latter field is always understood to be time independent.

The Principle of Equivalence. This principle is taken to encapsulate the ramifications of the equality of inertial and gravitational mass for gravitation theory. It has very many different forms, some of which include the equivalence of acceleration and gravitation. The form referred to in the text will be apparent in context. See Section 5.1 for further discussion.

The Principle of Relativity (of Special Relativity Theory) asserts the physical equivalence of all inertial frames of reference as far as the laws of nature are concerned. That is, no experiment aimed at discovering a law of nature can ever distinguish between inertial frames of reference and single out any as physically preferred.

The Principle of General Relativity asserts the physical equivalence of all frames of reference, as far as the laws of nature are concerned, where "physical equivalence" has the same import as above.

The Principle of General Covariance requires the formal equivalence of all spacetime coordinate systems. That is, the laws of nature must be capable of being written in a form which is the same in all spacetime coordinate systems.

The General Relativity of Motion asserts that the only meaning that can be ascribed to the motion of a body is that the body moves with respect to other reference bodies.

The Hypothesis of the Relativity of Inertia asserts that the inertial forces which act on an accelerating mass are produced by an interaction between that mass and all the remaining masses of the universe.

Mach's Principle asserts that the metric field is determined without residue by the masses of the universe and their distribution.

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

In 1907 Albert Einstein approached the task of incorporating gravitation into the framework of his new special theory of relativity. However, he soon found that this apparently straightforward task led to questions of far wider significance and import. It was not until eight years later, after some of the most intense and exhausting work of his life, that these questions were resolved. The theory which resulted, his general theory of relativity, is one of the most remarkable achievements of modern theoretical physics, a triumph of human speculative thought in its struggle to understand Nature.

We are very fortunate to have an extensive record of the steps of Einstein's struggle towards his final theory in the form of an almost continuous series of his journal articles and other secondary sources. In them we find a detailed account of the various stages of the developing theory and the organic network of concepts and ideas which surrounded them.

In this dissertation I analyse Einstein's development and use of this network of concepts and ideas and the theories which evolved from them. In particular I study the origin, content and ramifications of three of Einstein's most important heuristics: the belief that gravitation could not be dealt with adequately within special relativity, but that it required an extended theory of space and time; the equivalence of acceleration and gravitation; and the concept of the general relativity of motion.

The first of these is dealt with in Chapters 2 and 3. In Chapter 2, I examine the tradition of criticism of gravitation theory at the time of the emergence of the special theory of relativity, in order to display how the emergence of the theory served to combine, reinforce and redirect the disparate themes of that tradition. In Chapter 3 I analyse how Einstein came to his audacious 1907 conclusion that, in spite of this, his new-born special theory of relativity could not contain an acceptable gravitation theory. I try to reconstruct his attempts to erect such a theory and note that his rejection of such theories was based on their apparent failure to predict no influence by the sideways velocity of a freely falling body on its downward rate of fall. I argue that Einstein could have recovered such a result within a special relativistic gravitation theory if he had considered a second rank tensor theory.

In the next two chapters, I turn to the equivalence of acceleration and gravitation. In Chapter 4, I review the content and status of the principle of equivalence in the various stages of the developing theory. We shall see that it began as a specific postulate of the indistinguishability of the effects of uniform acceleration and homogeneous gravitational fields. In the years prior to 1913, it provided a bridge over which Einstein could pass from the accessible structure of inertial fields to the more obscure structure of gravitational fields, thus enabling the construction of a new relativistic gravitation theory. With the emergence of the first outline of the general theory of relativity in 1913, both the content and function of the principle in Einstein's formulations of the theory was to change, even though this was not explicitly recognised by Einstein until some five years later. Then he recognised that the significance of the equality of inertial and gravitational mass lay in the fact that the same theoretical machinery could be used to account for both inertial and gravitational phenomena. In the case of his theory, this meant that both were represented by the metric tensor. In 1913, Einstein also made a subtle but significant shift in his understanding of the empirical fact which underpinned the principle of equivalence. He no longer required that the downward acceleration of a falling body be independent of its sideways velocity. This requirement had been the basis of his 1907 rejection of the possibility of a special relativistic gravitation theory, but was not be entailed by his final general theory of relativity.

In Chapter 5, I examine the extent to which Einstein's final general theory of relativity will allow acceleration to stand for gravitation, as postulated by Einstein's original version of the principle of equivalence. Within this theory, inertial fields only mimic certain gross but crucial featues of gravitational fields, but do not mirror their finer structure. This accounts for Einstein's early success in recovering crucial features of his general theory of relativity, without the need to invoke that theory's sophisticated conceptual and theoretical apparatus. However it led him to conclude, in particular, that static gravitational fields would be spatially Euclidean, in contradiction to his final theory. This led to his recovery of only a "half deflection" for the gravitational bending of light by the sun. This conclusion was also carried over into the early 1913 versions of his general theory of relativity, where it played a role in the premature rejection of his final generally covariant field equations.

In Chapter 6, I review the concept which came to dominate Einstein's work on his theory after 1912, the general relativity of motion, and the critique of space which underpinned it. This critique demanded the final elimination of space as a causal entity from physical theory and led Einstein to an account of the origin of inertia as a gravitational interaction between the bodies of the universe. We shall see that this account of the origin of inertia was especially important to the early development of the general theory of relativity. The discovery of effects consistent with it in Einstein's earlier scalar theory of gravitation had mediated the emergence of the outline of the new theory in 1913 and had given Einstein confidence that his new theory was very close to a general theory of relativity, even though he had been unable to discover generally covariant field equations within it.

However, I argue that Einstein's critique was both ill-born and ill-fated. It had arisen through a misunderstanding of the import of Mach's critique of the Newtonian concept of space and was carried through in a corpuscular world view which saw the only causally active entities in the universe to be particles in motion in a passive and featureless container space. This view, in particular, was essential to the recovery of his account of the origin of inertia from his critique. The overthrow of this world view was only then being consummated by Einstein's own theory of relativity. This theory promoted the view that space had to be seen to be permeated with causally active fields. We shall see that Einstein even came to concede a view quite antithetical to his original critique: the general theory of relativity, in promoting the concept of the unified field, led to the view that space may well be the ultimate medium of all reality.

Thus I review how Einstein's original demand for the final elimination of space as a causal agent gave way to a weaker but related requirement - that space no longer be a cause that acts but is not acted upon itself. This found satisfactory expression in the "no prior geometry" feature of the final theory. The demand for the general relativity of motion and the physical equivalence of all frames of reference gave way to the requirement of the formal equivalence of all coordinate systems for the writing of the laws of nature and the insistence on generally covariant formulation of physical theories. Finally, we shall see that Einstein had to abandon his early ideas on the origin of inertia, in spite of intensive attempts to transfer them to the new field theoretic framework, although he never doubted the essential soundness of the ideas which gave birth to them.

In Chapter 7, I return to the examination of Einstein's theory as a relativistic theory of gravitation. I argue that the concepts and arguments which guided Einstein on his path towards his final theory - the principle of equivalence, his critique of space and the considerations which surrounded them - were volatile and at times contradictory. In spite of this I argue that we can still understand the discovery of the theory as the rational outcome of a single process if we recognise that, beneath Einstein's grand vision of a theory of the general relativity of motion, lay a more mundane objective, that of constructing a relativistically acceptible gravitation theory. This provided a constraint envelope, which could control and direct Einstein's work towards his final theory, and a stable framework to counter the volatility of the other considerations which guided him. I claim that if Einstein were to carry through the implications of this constraint envelope to the full, then this in itself would guarantee that he would arrive at his final general theory of relativity or one very much like it. I demonstrate how Einstein refused to comprise the requirements of this constraint envelope with the example of his derivation of non-generally covariant field equations in 1913 and his recognition in 1912 that the equivalence of uniform acceleration and gravitation could only hold locally in his 1912 scalar theory of gravitation.

In Chapter 8, I illustrate the power of the constraint envelope to lead to a theory of gravitation of the nature of Einstein's final general theory of relativity by a case study of the development of Nordström's scalar theory of gravitation in the 1912-1914 period. In particular I show how only the sustained application of consistency requirements similar to those of Einstein's constraint envelope were sufficient to lead to a final theory which burst the bounds of special relativity. In its most transparent form, the theory accounts for gravitation as the curvature of spacetime, as is the case in the general theory of relativity, and with a field equation based on the Riemann curvature tensor.

I also present and extend a remarkably compact but little known derivation of the theory given by Einstein in 1913. The derivation begins by trying to construct a scalar theory of gravitation within special relativity and clearly demonstrates how effects amounting to a breakdown of the spacetime theory of special relativity inexorably arise.

CHAPTER 2

GRAVITATION BEFORE RELATIVITY THEORY

2. GRAVITATION BEFORE RELATIVITY THEORY

Throughout the years of his search for a general theory of relativity, Einstein's attention was explicitly focussed on the problem of constructing a new gravitation theory and, in the early years of this search, almost exclusively so. Einstein's work on this problem came as a natural extension of the well established but hitherto unfruitful tradition of criticism of Newtonian gravitation theory. In this chapter, I outline the state of development of this tradition so that we may see later how Einstein's work came as a consummation of many of its essential themes and to enable us to appreciate the novelty of Einstein's approach and the originality of his contribution.

Finally we shall see that the research programmes associated with each of these themes of criticism, when taken individually, were neither compelling nor complete. Further, whilst their combined weight was formidable, it was undirected. They lacked a unifying theme, which could weld them together into a united and overwhelming force and which was soon to be provided.

2.1 Astronomical Anomalies

Empirically, Newtonian gravitation theory was undoubtably one of the most successful of all physical theories. But, by 1900, it had become apparent that there were a number of small but persistent deviations from its predictions in the motion of celestial bodies. These became the focus of a tradition of criticism of the then current understanding of the structure of the planetary system and, further, of Newtonian gravitation theory itself. [1]

Zenneck, in his 1901 review article of gravitation, was able to draw a list of six anomalies from the work of Newcomb. [2] The best known of these was the anomalous advance of the perihelion of Mercury of about 40 seconds of arc per century. Also deviations in the motions of the nodes of Venus and the perihelion of Mars were five times and three times in excess of the probable error, respectively. Further, it was suspected that the eccentricity of Mercury differed from the expected by more than twice the probable error. Finally, there were anomalies in the motion of Encke's comet and small irregularities in the motion of the moon.

A large number of attempts were made to account for these anomalies within the domain of Newtonian theory. [3] After Leverrier discovered the anomalous motion of Mercury, he proposed that it might be due to the perturbations caused by an as yet undiscovered planet in the region of Mercury. This hypothesis failed mainly

---

1. S. Newcomb, The Elements of the Four Inner Planets and the Fundamental Constants of Astronomy. Supplement to the American Ephemeris and Nautical Almanac for 1897 (Washington: Government Printing Office, 1895); J. Zenneck, "Gravitation", Encyklopädie der Mathematischen Wissenschaften mit Einschluss ihrer Anwendungen (Leipzig; B.G. Teubner, 1903-1921: concluded in Aug. 1901), pp.25-67; S. Oppenheim, "Kritik des Newtonschen Gravitationsgesetzes", Encyklopädie der Mathematischen Wissenschaften mit Einschluss ihrer Anwendungen (Leipzig: B.G. Teubner, 1922-34: concluded in Feb. 1920), pp.80-158; J. Chazy, Théorie de la Relativité et la Mecanique Céleste (Paris: Gauthier-Villars, 1928); L. Pyenson, "The Goettingen Reception of Einstein's General Theory of Relativity", Diss. Johns Hopkins 1974.

2. Zenneck, p.36; Newcomb.

3. See especially Oppenheim, pp.132-46; Newcomb, pp.111-8.

because the astronomers were unable to find such a planet. However, the hypothesis that the anomalies were due to hitherto undiscovered matter distributions was retained and reappeared in some of the most influential explanations.

The locations of this extra matter varied. One hypothesis placed it in an undiscovered moon of Mercury. Another suggested that the sun was not perfectly spherical. Both of these were soon regarded to be unacceptable. [4] The most promising approach lay in the hypothesis that the extra matter was to be found in belts of interplanetary planetoids or even in dust clouds which were hypothesised to comprise the Zodiacal light. Newcomb, in 1895, rejected these hypotheses as untenable. They would either produce an insufficient effect or cause unobserved perturbations in the motion of some planets. [5]

In 1906, the respected astronomer Seeliger revived and revamped the Zodiacal light hypothesis. In this form the hypothesis was to last until the early 1920's. In 1920, Einstein looked upon the possibility of a distribution of interplanetary matter as the only alternative to his general theory of relativity when it came to accounting for the anomalous motion of Mercury's perihelion. [6] The majority of astronomers, however, approached the hypothesis with caution. [7]

---

4. Oppenheim, pp.134-6; Newcomb, pp.111-2.

5. Newcomb, pp.112-8.

6. A. Einstein, "Meine Antwort auf die antirelativitätstheoretische G.m.b.H", Berliner Tageblatt und Handelszeitung, 27 August, 1920 pp.1-2.

A number of other hypotheses were suggested to account for the various anomalies whilst remaining within the scope of Newtonian theory. To account for the deviations in the motion of Encke's comet, Encke had suggested that its motion was impeded by a resisting medium. This in turn might also be seen to be due to the matter of the Zodiacal light, as was later suggested. The possibility of light pressure from the sun causing perturbations was also considered. Another source of anomalies in the motion of heavenly bodies that was considered was the possibility of changes in the rotational period of the earth or through tidal actions. Finally, the effect of an increase in the mass of the earth and heavenly bodies due to matter collected in their passage through the matter supposed to comprise the Zodiacal light was considered. All these hypotheses met with only limited success, however.[8]

The anomalies could also be accounted for with some success by allowing small modifications to Newton's law of gravitation. This is the method that Newcomb favoured. Specifically he accounted for the anomalous motion of the planets in terms of an hypothesis due to Hall, who had suggested that gravitational force does not fall off with the square of distance, but with distance raised to the power $2+\delta$, where Newcomb set $\delta = 0.000\,000\,1574$.[9]

---

7. J.D. North, The Measure of the Universe (Oxford: Clarendon, 1965), p.48; Oppenheim, pp.136-9.

8. Oppenheim, pp.139-46.

9. Newcomb, pp.119-23.

Zenneck reviewed the success of a number of modifications of Newton's law with regard to their accounting for the anomalies. He looked mainly at those which were based on the assumption of a finite velocity of propagation for gravitation or, as a special case of this, those that were suggested through analogies with electro-magnetism. Whilst not all were successful, some worked reasonably well and sufficiently so for Zenneck to conclude that Newton's law could now only be seen to hold for stationary bodies and that the assumption that gravitational action propagated at the speed of light could remove the tiniest differences between observation and theory. [10]

Thus we can see that these anomalies had become an integral part of the developing tradition of criticism of Newtonian gravitation theory. Indeed they had become one of the most powerful tests of any new gravitation theory. However, especially in view of the strongly empiricist leanings of modern thinking, we must not over-estimate their importance in this developing tradition and their ability to support a comprehensive programme of modification of Newtonian theory.

It must be recalled that Newtonian gravitation theory had been one of the most successful physical theories of all time. On its shoulders rested the new world view which had emerged as the resolution of the Copernican revolution. It had managed to overcome many apparent observational contradictions. For example, there was Laplace's solution of the great inequality of Jupiter and Saturn. [11]

---

10. Zenneck, pp.46-51

11. E. Whittaker, A History of the Theories of Aether and Electricity Vol.II. The Modern Theories 1900-1926 (London: Nelson, 1953), pp.144-5.

Most notable in then recent memory was the discovery of the planet Neptune in 1846, a discovery which resulted solely from the assumption that an anomalous perturbation in the orbit of Uranus must be due to the existence of another body, in accord with Newtonian theory.[12]

Clearly, in the face of such a successful history, any attempt at modifying Newtonian theory which was based solely on the discovery of a few minute astronomical anomalies would need to be approached with caution.

Thus, in his review, Zenneck could write

> "Therefore small corrections to Newton's law are not excluded on the basis of astronomical experience, if also - especially in the cases denoted by 5 and 6 [anomalies in the motion of Encke's comet and small irregularities in the moon's orbit], in which the relationships are more complicated and more uncertain than with the planetary orbits - it is in no way settled, that the given differences have their basis in an inexactness of the gravitation law." [13]

Whatever approach one took to the significance of the anomalies, one was still faced with the same basic problem: the minute size of the deviations made their significance quite ambiguous. If one assumed that they could be explained within Newtonian theory by some as yet undiscovered feature of the planetary system, it was conceivable that very many different hypotheses could be applicable.

---

12. A. Berry, A Short History of Astronomy (1898; New York: Dover, 1961), pp.371-2.

13. "Kleine Korrektionen am Newton'schen Gesetz sind also auf Grund der astronomischen Erfahrung nicht ausgeschlossen, wenn auch - inbesondere in den unter 5 and 6 aufgeführten Fällen, in denen die Verhältnisse komplizierter und unsicherer liegen als bei den Planetenbahnen - keineswegens ausgemacht ist, dass die angegebenen Differenzen in einer Ungenauigkeit des Gravitationsgesetzes ihren Grund haben."

    Zenneck, p.36, his emphasis.

Similarily if one looked to a modification of Newtonian theory, the number of acceptable modifications would be conceivably very great. Indeed the number of ad hoc corrections to Newtonian theory which could account for the anomalies was only limited by one's skill at curve fitting.

As a basis for the development of new gravitation theories, the anomalies were simply insufficient. They were consistent with far too many theories and hypotheses - although finding them might well be a difficult task. The discovery of one that worked was no guarantee that it was the correct one.

The basis of a new gravitation theory would have to lie elsewhere - in strong theoretical or philosophical considerations. Only then could one turn to the anomalies as a test for the new theory. Even then one was not forced to reject the theory if the anomalies were not completely accounted for. For example, Lorentz discovered that his 1900 electromagnetic theory of gravitation could not explain the anomalous motion of Mercury and he concluded:

> "... if we do not pretend to explain this inequality by an alteration of the law of attraction, there is nothing against the proposed formulae." [14]

Yet in the same paper, he rejected his earlier electromagnetic theory of gravitation principally on the grounds of its unsafisfactory energetics. [15]

Thus it appears that for Lorentz theoretical considerations - in this case energetics - and philosophical considerations - in this

---

14. H. A. Lorentz, "Considerations on Gravitation", Proceedings of The Royal Academy of Sciences, Amsterdam, 2 (1900), 573.

15. Ibid, p.565.

case that gravitation should be reducible to electromagnetism - were more important than the empirical question of the anomalies.

I now turn to look at the types of philosphical and theoretical considerations that guided attempts to formulate new gravitation theories at the time of the birth of special relaivity.

2.2 Reductionism

One of the most persistent themes in the tradition of criticism of gravitation theory since the time of Newton has been the continued attempts to reduce gravitation to some other phenomenon. In general, underlying such attempts is a belief in the possibility of a simple and unified account of all phenomena in terms of a very few basic processes. In the cases considered, Newtonian gravitation did not qualify as one of these basic processes, perhaps because of its action at a distance nature. I examine two types of reduction: mechanical and electromagnetic.

Mechanical Theories [16]

In keeping with the mechanistic world view, mechanical explanations of gravitation have been popular since the time of Newton. Such theories have been attributed to notables including Newton,

---

16. See Zenneck, pp.53-64; North, pp.32-40; Pyenson, pp.25-38; W. Ritz, "Die Gravitation", Rivista di Scienza, 5 (1909), pp.241-255; M. Jammer, Concepts of Force (Cambridge, Mass: Harvard U.P., 1957), pp.135-141, 188-199.

Hooke and J. Bernoulli. [17]

In the nineteenth century, mechanical theories enjoyed considerable popularity, although they had mostly fallen from favour by the end of the century. There were a number of different strands in these theories. One was founded by Bjerknes in the 1870's on the basis of some unusual results in hydrodynamics. It was found possible to recover an inverse square law of attraction between two bodies if they were to vibrate in phase and be contained in a medium which would transmit the vibrations of one body to another. The theory was modified by later workers to save it from a number of objections, but nonetheless did not survive into the twentieth century. [18]

Another strand was linked into the ether theories of matter, which were especially popular in England in the latter half of the last century. In these theories, matter atoms were seen as vortices or sinks and sources in the ether and an inverse square attraction law was extracted from the properties of the aether. [19]

Another line of theorising which attracted much discussion was based on Thomson's 1872 revival and modification of an older theory due to leSage. In leSage's theory, bodies were seen to be immersed in a sea of rapidly moving "ultramundane corpuscles" with which they constantly collided. When two bodies approached, they would shield each other from the impact of the corpuscles in such a way as to

---

17. Jammer, loc.cit.

18. North, pp.32-4; Zenneck, pp.54-7.

19. North, pp.34-8; Zenneck, pp.53-4.

cause an attractive force consistent with the inverse square law. Thomson and later workers tried to modify this basic idea in order to save the theory from many objections, yet they were ultimately unsuccessful. In his review article, Zenneck could still list six major objections to the theory. The first dealt with the inevitable loss of energy which accompanies the inelastic collisions of the corpuscles with bodies. Other objections question the self-consistency of the theory and its consistency with known phenomena.[20]

By the end of the century, mechanical theories had fallen from favour, especially outside England. First, this came about because of the nature of the theories themselves. They usually required a number of quite ad hoc and untestable hypotheses about complex mechanisms supposedly underlying the phenomena and often involved densities of energy far in excess of that seen to be involved in the phenomena directly. Second, there was an important shift in the world view of many physicists away from a mechanical view to a field theoretic and, more specifically, electromagnetic view with the great steps forward taken in electromagnetic theory.

Electromagnetic Theories [21]

The growing strength of electromagnetic theory in the nineteenth century was reflected in gravitation theory. First there were attempts simply to use the mathematical form of electromagnetic laws

---

20. North, pp.38-40; Zenneck, pp.57-64.

21. See Zenneck, pp.46-8, 65-7; North, pp.46-8; Pyenson, pp.41-58.

in a somewhat ad hoc manner as laws of gravitation.[22] Such attempts were not reductions in the usual sense. Such reductions were soon to follow, however.

The best known of these was Lorentz's 1900 theory. Lorentz had first considered a theory equivalent to the mechanical theories based on leSage's ultramundane corpuscles, but in which the pressure of corpuscle impact was replaced by electromagnetic radiation pressure. He gave up this theory because it required the continued vanishing of energy and was also observationally unsound. Lorentz then turned to a theory based on an idea used by Mosotti, Zöllner and Weber. The theory accounted for gravitation as the result of the force of attraction between unlike charges being slightly greater than the force of repulsion between like charges, for equal quantities of charge. The theory could not, however, account for known astronomical anomalies.[23]

At the time of the birth of special relativity, electromagnetic reductions seemed an unproven but promising approach and a number of other such theories were being developed at that time.[24]

Discussion

Reductionism has met with a number of striking successes, such as the case of the kinetic theory of gases. In such cases, the reduction enables a far-reaching simplification of unification of our understanding of the basic processes of nature. However, in

---

22. Zenneck, pp.46-8.

23. Lorentz, op.cit.

24. Pyenson, pp.41-58; Ritz op.cit.

the case of gravitation, no such simplification had yet been achieved by means of a reduction. Indeed mechanical reductions, which had fallen out of favour, seemed to complicate the world view rather than simplify it, whilst electromagnetic reductions remained unproven. Few reductions could compete with the simplicity of Newton's inverse square law - although with it one had to swallow instantaneous and unmediated action at a distance.

So Zenneck ended his review article with the conclusion that all attempts to link gravitation in a satisfactory way with other areas of experience had either failed or were uncertain. He saw the basic understanding of gravitation at the start of the twentieth century as no further advanced than at the start of the eighteenth century.[25]

2.3 Cosmological Difficulties for Newtonian Gravitation Theory[26]

A cosmological difficulty for Newtonian theory was discussed by Neumann in 1874 and elaborated by Seeliger in 1895. If one assumed that the universe was spatially infinite and contained a fairly uniform matter distribution, then it followed within Newtonian

25. Zenneck, p.67.

26. See Zenneck, pp.51-3; Oppenheim, pp.148-151; North, pp.17-8; Pyenson, pp.18-21; M. Jammer, Concepts of Mass (Cambridge, Mass: Harvard U.P., 1961), pp.127-31.

theory that the gravitational force exerted by the masses of the universe on a test mass at any given point in space would be mathematically indeterminate.

Seeliger proposed a resolution of the difficulty which involved the addition of an exponential decay factor to the usual Newtonian force law. This factor could be interpreted as corresponding to a minute absorption of gravitational force in its transit through space. A similar modifications of the Newtonian potential formula, originally proposed by Neumann in another context, was considered as well as the possibility that the distance exponent of the inverse square law might be slightly greater than two. This latter suggestion had been made by Hall as an explanation of the anomalous motion of Mercury.

However, each of these attempts were seen to be unsuccessful. If the exponential decay factors were used, they could be adjusted to account for only some but not all of the know astronomical anomalies. If the exponent of the inverse square law was adjusted to account for these anomalies, it could no longer solve the cosmological problem. [27]

However, this cosmological objection is not adequate as a basis of a programme of modification of Newtonian theory. First, the objection is consistent with far too many possible hypotheses. There are, as we have seen, quite a few simple modifications that

---

27. Zenneck, pp.51-2; Oppenheim, pp.148-151

satisfy the demands of this particular objection although they may be rejected on other grounds.

Second, and of more importance, is the fact that modifications to gravitation theory are based on a particular cosmology, which presumed a spatially infinite universe with a uniform matter distribution. Whilst Newtonian gravitation theory was one of the most thoroughly tested and corroborated of all theories, the cosmology that was being advanced was untested, perhaps incapable of proof and only supported by aesthetic preference. A small change in this cosmology might be all that was necessary to radically change or even negate the entire critique.

For example, Föppl suggested that the problem could be resolved without modifying Newton's law, if one assumed that, in addition to normal matter, the universe contained an equal quantity of "negative" matter. Such matter, it was suggested, would be repelled by normal matter, since it has an opposite gravitations "charge". Although the idea was not without its own problems, it attracted quite some attention.[28] For our purposes, however, this example shows the fragility of any critique of Newtonian theory, based on such a cosmology. Certainly, a far safer approach is to work from a trusted physical theory to a new cosmology as Einstein was to do several years later, rather than the other way round.[29]

---

28. Zenneck, pp.52-3; Jammer, Concepts of Mass, pp.128-131.

29. See Chapter 6.

2.4 Gravitation and Field Theory

The Concept of the Field [30]

The nineteenth century saw the introduction of a concept which has been described as "...one of the most significant philosophical and scientific steps in the past several centuries...", the introduction of the concept of the field. [31] Historically, the origins of the concept lie in the belief that separated bodies cannot interact without some form of mediation. This belief was already well established at the time of Newton and is reflected in the reception of Newton's theory of gravitation, a theory which appeared to propose instantaneous and unmediated action at a distance. Certainly, the general discomfort felt over a theory based on such action was reflected in the persistent attempts aimed at reducing gravitation to some sort of mechanical phenomenon, attempts which were popular from the time of Newton well into the last century.

It was not until the nineteenth century that this discomfort gave birth to the concept of the field in the study of electricity and magnetism. From the work of Faraday and Maxwell it became apparent that the most satisfactory account of electricity and magnetism was to be found with the use of the field concept, a concept which had been developed for exactly this purpose.

---

30. See M. Hesse, Forces and Fields (Totowa, N.J.: Littlefield, Adams & Co.), 1965; W. Berkson, Fields of Force (London: Routledge and Kegan Paul, 1974); North, pp.25-32.

31. M. Sachs, The Field Concept in Contemporary Science (Springfield, Ill.: Charles C. Thomas, 1973), p.vii.

Gravitation, however, stubbornly refused to be drawn into the realm of field theory. In retrospect we can understand this by noting that gravitation was not known with certainty to satisfy any of a number of requirements which we might now delineate for judging the physical reality of a field. Such criteria may require the field to be independently detectable; to propagate action at a finite speed; or for this propagation to be affected by material changes in the intervening medium; or they may require the field to admit a mechanical model or for momentum and energy of the interactions it mediates to be capable of being located in the intervening space.[32]

In spite of this, gravitation, as an unmediated action at a distance theory was approached with much suspicion in many quarters. Faraday, for example, felt that the Newtonian action at a distance theory violated his requirement of the conservation of "force", a concept which has no clear meaning in modern theoretical terms but is composed of a hybrid of the modern concepts of force and energy. In particular, when two masses approached under the action of their mutual attractive forces, these forces would increase in accord with the inverse square law. Farady asked where could this force come from? It was not apparently impressed by some outside agent. It did not come from the removal of gravitational force from some other body to be concentrated on the approaching body. Nor could Faraday trace its origin experimentally to the force depleted from some other appropriate phenomena, such as an electrical phenomenon.

---

32. Hesse, p.197, p.204.

Or was it that the force was always distributed in space but only made itself apparent when another body approached? [33]

But a field theory of gravitation, in which the field was recognised as physically real, could not be found. Up to the time of Faraday, as we have seen, the most popular attempts involved mechanical or hydrodynamical field theories. It is interesting to note the difficulties Maxwell encountered in his attempts to produce a mechanical ether theory of gravitation. They follow directly from Faraday's observation on the conservation of "force".

Since gravitational forces between masses - bodies of like gravitational "charge" - are attractive, it follows that the energy density of the gravitational field in the neighbourhood of a heavy body is less than that in regions distant from that and other bodies. This is the opposite of the case of the field of an electric charge. Since like charges repel, the energy density of the electric field in the neighbourhood of a charge is greater than in regions remote from it and other charges. Now Maxwell required the energy density of the field to be positive everywhere. This means that in a region where the net gravitational force on a test body is zero, for example in a region infinitely remote from other bodies or at the gravitational null point between the earth and sun, the field energy density must have an enormous value - large enough to ensure that the energy density of the strongest possible field in the entire universe will still be positive. This was sufficient to lead Maxwell to give up this line of research, for he could not understand how a medium could possess such properties. [34]

---

33. Hesse, pp.22-4; Berkson, pp.112-6.

34. North, pp.30-31; Hesse, pp.224-5.

Maxwell's problem is not merely an artefact of the particular style of field theory he chose to adopt. It is one that confronts all Newtonian gravitational field theories. We can carry it further. On the basis of these considerations we would expect that a body collapsing gravitationally onto itself could yield an unbounded quantity of energy as the collapsing body nears the formation of a point mass - and there is nothing in a simple Newtonian gravitation theory to say that it cannot collapse arbitrarily close to a point.

This difficulty was only resolved after the advent of relativity theory in which the equivalence of mass and energy enabled an alternate interpretation of the energetics of gravitational collapse.

Speed of Propagation [35]

A question which is intimately linked to the problem of the reality of the gravitational field is the question of the speed of propagation of gravitational action. Indeed if it can be shown that gravitational action propagates at a finite speed, then one has strong grounds for arguing for the reality of the gravitational field. The converse - that the reality of the field leads to the conclusion that the action must propagate at finite speed - does not follow, as Hesse has pointed out. [36] A mechanically rigid medium can propagate actions at infinite speed in Newtonian theory, for

---

35. See Zenneck, pp.46-51; Oppenheim, pp.152-158; North, pp.44-48; Pyenson, pp.15-18.

36. Hesse, p.198.

example. However, if one was commited to the reality of the field and the rejection of unmediated action at a distance, then one would seek to find evidence of a finite speed of propagation as the most unequivocal evidence available, perhaps recognising that a failure to find such a finite speed would be awkward but not fatal.

Oppenheim has divided the many attempts to incorporate a finite speed of propagation into gravitation theory into two groups: those that involve supplementary terms in (v/c) and those in $(v/c)^2$, where v is the speed of a body such as a planet and c is the speed of propagation.[37] The most famous of the first type of attempts was Laplace's. From the secular acceleration of the moon, Laplace was able to put a lower limit on the speed of propagation of 7×10^6 of the speed of light. This clearly justified the assumption that this speed could be taken to be infinite for all practical purposes. However, theories of this first type met little success in trying to account for the known astronomical anomalies.

The second group generally involved the use of an electrodynamic analogy. So, for example, the forms of electrodynamic action at a distance laws of Weber and Riemann were used as gravitational laws. This naturally involved the assumption that the speed of propagation of gravitation was equal to that of light, although one was not forced to take this value.

These attempts were more successful than the first group, some being able to account for well known astronomical anomalies. Lévy and Gerber, both working in the 1890's, were able to account for the

---

37. Oppenheim, p.152.

the anomalous motion of Mercury's perihelion. These attempts were, however, somewhat ad hoc. For example, neither Weber's nor Riemann's electrodynamic laws could individually account for the motion of Mercury's perihelion, when transplanted into gravitation theory. Lévy, however, simply took an ad hoc combination of the two laws, the proportions of each determined solely by the requirement that Mercury's motion be accounted for.

This shows that the idea of finite speed of propagation was considered very seriously in its own right at the time of the birth of special relativity. It should be noted, of course, that all the reductions of gravitation which have already been discussed - the mechanical, hydrodynamical and electromagnetic - automatically involve the assumption of a finite speed of propagation. But the proliferation of such theories shows that it was by no means clear how this velocity was to be incorporated into gravitation theory.

Discussion

By the end of the last century there was considerable discomfort with the fact that Newtonian gravitation theory was based on unmediated action at a distance. It was expected, contrary to Newtonian theory, that gravitational action propagated at a finite speed and that this would enable an explanation for some of the known astronomical anomalies. Further, it was felt that gravitation should be represented by a field theory. This manifested itself in the attempts to account for gravitation by reduction to other field-like phenomena such as electrodynamics and hydrodynamics, as well as through explicit attempts to formulate a gravitational field theory.

However, no clear picture had emerged of exactly how a finite speed of propagation was to alter the structure of Newtonian gravitation theory. Further, the problem of reformulating Newtonian gravitation theory as a true field theory was by no means solved. A trivial rewriting of Newtonian theory using the machinery of field theory was no problem. In his review, Zenneck noted that a field described by

$$\text{rot}\ \underline{R} = 0 \qquad \text{and} \qquad \text{div}\ \underline{R} = 0 \ \text{ or } = -4\pi G\rho \ \text{(if matter is present)},$$

were $\underline{R}$ is the gravitational field strength, ρ the density of matter and G the gravitational constant, is identical to Newtonian theory. This, he noted, offers no advantages or disadvantages over traditional Newtonian theory. [38]

Of course this theory is really a field theory in name only. Within it, the gravitational field is an entirely dispensible entity. In particular it satisfies none of the requirements which we might make for the physical reality of a field. Propagation of gravitational action is instantaneous, it is unaffected by the state of the intervening space and there appears to be no way to independently detect the field.

Zenneck continued to note that, if the speed of propagation of gravitation was finite, then the field approach could offer advantages, with Newtonian theory being naturally extended along the lines of electromagnetism. He concluded, however, that this approach would be troubled by Maxwell's problem of the field energy density, which I discussed above.

---

38. Zenneck, pp.64-5

2.5 Conclusion

By the end of the last century, it was felt that gravitation theory had begun to stagnate and that it had not made any fundamental advances since the seventeenth century. This was in contrast to other branches of science and , in particular, electromagnetic theory. Indeed, gravitation theory had been used as a model to assist in the construction of theories of electricity and magnetism; now electromagnetic theory was being used to suggest modifications to gravitation theory.

Further, there had been a number of developments which highlighted the stagnation of gravitation theory: the establishment of the field concept, the emergence of the electromagnetic world view and the persistence of a number of small astronomical anomalies. It seemed that gravitation theory could accommodate none of these in a universally satisfying manner.

As we have seen, there were several programmes of modification of Newtonian gravitation theory, which were based on these and other considerations. However, each programme taken individually seemed incomplete for one reason or another. In particular, no single programme could insist on the necessity of a modification to Newtonian theory. For example, one could always reply that the astronomical anomalies could be explained by the distribution of undiscovered matter in the planetary system, or that the belief that gravitation can be reduced to electromagnetism was unfounded, or that no matter how aesthetically displeasing, unmediated action at a distance did occur and that the success of Newtonian theory had proved it.

If, however, all these programmes were taken together, they combined into a formidable but undirected case against Newtonian gravitation theory. What was still missing was a key which could bring the elements of each programme together to form a coherent whole, display the necessity of a modification of Newtonian gravitation theory and provide solid guidelines for the construction of such a modification.

This key was soon to emerge. It emerged as one of the most important consequences of the field concept and the new discoveries in electricity and magnetism. It emerged as the resolution of the conflict between the old Newtonian world view and the new world view promoted by these discoveries. It was the theory of relativity.

Relativity theory, initially in its special form, combined and reinforced central ideas from the tradition of criticism of Newtonian gravitation theory. It guaranteed that the accuracy of Newtonian gravitation theory had to be limited and that small deviations from its predictions were to be expected at the boundaries of the observable. It insisted that there was an upper limit to the speed of propagation of gravitation, as for any process, and gave its value as the speed of light in a vacuum. In so doing, it ensured that a field account of gravitation would have to emerge, for a true field would now have to be introduced into gravitation theory to mediate the interactions of separated masses. Also relativity theory teased out a new understanding of the nature of matter and energy, which had been foreshadowed by the results of electromagnetic theory and would now have to appear in a relativistically acceptable gravitation theory. Finally, the gravitation theory which was to emerge from this over a decade later, Einstein's general theory of relativity, was to explain many of the outstanding astronomical anomalies, most notably the anomalous motion of Mercury, without the need to resort to any ad hoc hypotheses. Further, the theory proved itself far more capable than Newtonian theory of producing consistent and interesting cosmologies - so much so that it fostered a widespread rebirth in the study of cosmology, unparalleled since the time of Newton himself.

The emergence of the special theory of relativity at last provided a solid theoretical vindication of the necessity for a modification of Newtonian gravitation theory. Of greater significance, we shall see in the next chapter that special relativity theory provided comprehensive guidelines for the form that the new theory should take. Thus special relativity theory combined and eclipsed the themes of the tradition of criticism of Newtonian gravitation theory. It gave the search for a new gravitation theory a definite direction and reduced it to the single task of constructing a relativistically acceptable gravitation theory.

CHAPTER 3

THE FIRST RELATIVISTIC GRAVITATION THEORIES

3. THE FIRST RELATIVISTIC GRAVITATION THEORIES

The emergence of the special theory of relativity crystallised and reoriented the established tradition of criticism of Newtonian gravitation theory. It provided a solid theoretical vindication of the belief that Newtonian theory was in need of modification and, as we shall see, gave clear guidelines on the form which these modifications must take.

Einstein was one of a number of physicists who worked on the problem of developing a relativistically acceptable gravitation theory at the time of the emergence of the special theory of relativity. Poincaré and Minkowski also worked on this same problem. But it was Einstein who first came to recognise that the construction of a relativistically acceptable gravitation theory involved far more than a simple mechanical exercise in the application of the techniques of relativity theory. He was to conclude within two years of his famous 1905 paper on special relativity that attempts to account for gravitation within special relativity could not succeed. Gravitation was to burst out of special relativity.

In this chapter I analyse Einstein's understanding of the immediate impact of special relativity on gravitation theory and his earliest work in the area and contrast it with the corresponding work of Poincaré and Minkowski.

3.1 Einstein on the Impact of Special Relativity on Gravitation

In 1913 the revolution in gravitation theory precipitated by the advent of special relativity was in full swing. In that year, in the introduction to an address in Vienna on an early version of his general theory of relativity, Einstein gave a most revealing sketch of what he saw to be the impact of special relativity on gravitation theory. I quote this sketch, which is perhaps unique in Einstein's writings, at length. For it confirms that my account of the problem of gravitation and the impact of special relativity upon it follows Einstein's own perceptions of the matter at that time. Here Einstein also restates the problem, which is central to the concerns of this chapter, in an especially simple form and this will be followed up in the next section.

> "The region of phenomena of physics, of which there was first successful thoeretical illumination, was that of general mass attraction. The laws of gravity and the motion of the heavenly bodies were reduced by Newton to a simple law of the motion of point masses and to a law of the interaction of two gravitating point masses. These laws have proved themselves exactly right in such a way that from the standpoint of experience there is no decisive reason to doubt their strict validity. If, in spite of this, currently hardly more than one physicist might be found who believes in the exact validity of those laws, then this is to be traced back to the transforming influence which the development of our knowledge of electromagnetic processes has brought with itself in the last decade.
>
> "For, before Maxwell, electromagnetic processes had been traced back to elementary laws, which had been constructed as accurately as possible after the model of the Newtonian force law. According to these laws, electric masses, magnetic masses, current elements <u>etc</u>. should have exerted an action at a distance on oneanother, which requires no time for its propagation through space. Then H. Hertz showed 25 years ago through his ingenious experimental investigation on the propagation of electrical force, that electrical action requires time for its propagation. Thus he helped Maxwell's theory to the victory, which put partial differential equations in the place of unmediated action at a distance. Thus, after the unnacceptability of action at a distance theories had been proved in the region of electrodynamics, confidence in the correctness of Newton's action at a distance theory of gravitation was shaken.

This must have broken the path for the conviction that Newton's gravitation law spanned the phenomena of gravitation in their totality just as little as Coulomb's law of electrostatics and magnetostatics the totality of electromagnetic processes. That Newton's law was adequate until now for the calculation of the motion of heavenly bodies can be traced back to the fact that the speeds and accelerations of those motions are small. In fact it is easy to show that heavenly bodies whose motions would be determined through electrical forces, which orginate in electrical charges located on the heavenly bodies, would not reveal Maxwell's laws of electrodynamics to us, if the speeds and accelerations of those heavenly bodies were to be of the same order of magnitude as the motion of the heavenly bodies known to us. One would be able to represent those motions with great accuracy under the postulation of Coulomb's law.

"With this, confidence in the Newtonian action at a distance law was shaken. However, for the time being, there were no direct reasons to force the extension of Newtonian theory. But today there is such a reason for those who maintain the correctness of the theory of relativity. For, according to the theory of relativity, there are no means in nature which would allow us to send signals with a speed over that of light. But on the other hand it is obvious that with the strict validity of Newton's law we could employ gravitation to send signals instantaneously from a place A to a distant place B; for the motion of a gravitating mass in A would have to lead to simultaneous changes in the gravitational field in B - in contradication with the theory of relativity.

"But the theory of relativity brings us not only the compulsion to modify Newton's theory; fortunately it limits the possibilities for this in far reaching measure. If this were not the case, then the endeavour to generalise Newton's theory would have been a hopeless undertaking. In order to see this distinctly, let us transfer ourselves only into the following analogous situation: suppose that there be known of electromagnetic phenomena only what is known experimentally of those of electrostatics. One would know, however, that electric actions cannot propagate with speeds faster than light. Who would have been able to develop Maxwell's theory of electromagnetic processes out of this data? Our knowledge in the region of gravitation however corresponds exactly to the previous ficticious case; we only know the interaction between resting masses, and this apparently only to the first approximation. The confusing variety of possible generalisations is limited through relativity theory, in that following this theory the time coordinate enters in all systems of equations in the same way as the three space coordinates up to certain differences of sign. This great formal understanding of Minkowski's, which has been indicated here only inexactly, proved itself to be of greatest significance as an aid in the search for equations in accord with the theory of relativity." [1]

"Das Erscheinungsgebiet der Physik, dessen theoretische Durchleuchtung zuerst gelang, war dasjenige der allegemeinen Massenanziehung. Die Gesetze der Schwere und der Bewegung der Himmelskörper wurden von Newton auf ein einfaches Gesetz der Bewegung des Massenpunktes und auf ein Gesetz der Wechselwirkung zweier gravitierender Massenpunkte reduziert. Diese Gesetze haben sich als derart exakt zutreffend erwiesen, dass vom Standpunkte der Erfahurung aus kein entscheidender Grund vorliegt, an der strengen Gültigkeit derselben zu zweifeln. Wenn trotzdem gegenwärtig kaum mehr als ein Physiker sich finden lassen dürfte, der an die exakte Gültigkeit jener Gesetze glaubt, so ist dies auf den umgestaltenden Einfluss zurückzuführen, den die Entwicklung unserer Kenntnisse von den elektromagnetischen Vorgängen in den letzten Jahrzehnten mit sich gebracht hat.

"Vor Maxwell wurden die elektromagnetischen Vorgänge nämlich auf Elementargesetze zurückgeführt, die möglichst genau nach dem Muster des Newtonschen Kraftgesetzes gebaut waren. Nach diesen Gesetzen sollten elektrische Massen, magnetische Massen, Stromelemente usw. Fernwirkungen aufeinander ausüben, die zu iherer Fortpflanzung durch den Raum keine Zeit brauchen. Da zeigte H. Hertz vor25 Jahren durch seine geniale Experimentaluntersuchung über die Ausbreitung der elektrischen Kraft, das die elektrischen Wirkungen zu ihrer Ausbreitung Zeit brauchen. Er verhalf so der Theorie Maxwells zum Siege, die on die Stelle der unvermittelten Fernwirkung partielle Differentialgleichungen setze. Nachdem so auf dem Gebiete der Elektrodynamik die Unhaltbarkeit der Fernwirkungstheorie erwiesen war, war auch das Vertrauen in die Richtigkeit von Newtons Fernwirkungstheorie der Gravitation erschüttert. Es musste sich die überzeugung Bahn brechen, dass Newtons Gravitationsgesetz ebensowenig die Erscheinungen der Gravitation in ihrer Gesamtheit umspanne wie das Coulombsche Gesetz der Elektrostatik und Magnetostatik die Gesamtheit der elektromagnetischen Vorgänge. Dass Newtons Gesetz zur Berechnung der Bewegungen der Himmelskörper bisher hinreichte, ist darauf zurüchzuführen, dass die Geschwindigkeiten und Beschleunigungen jener Bewegungen klein sind. In der Tat ist leicht zu ziegen, dass Himmelskörper, deren Bewegungen durch electrische Kräfte bestimmt wären, die von auf den Himmelskörpern sitzenden elektrischen Ladungen herrührten, uns die Maxwellschen Gesetze der Elektrodynamik nicht verraten würden, falls die Geschwindigkeiten und Beschleunigungen jener Himmelskörper von derselben Grössenordnung wären wie bei den Bewegungen der uns bekannten Himmelskörper. Man würde unter Zugrundelegung des

Coulombsches Gesetzes jene Bewegungen mit grosser Genauigkeit darstellen können.

"War damit das Vertrauen auf die umfassende Bedeutung des Newtonschen Fernwirkungsgesetzes erschüttert, so lagen zunächst doch noch keine direkten Gründe vor, die zur Erweiterung der Newtonschen Theorie Zwangen. Ein solcher direkter Grund liegt aber heute für diejenigen vor, die an der Richtigkeit der Relativitätstheorie festhalten. Nach der Relativitätstheorie gibt es nämlich in der Natur kein Mittel, das uns gestatten würde, Signale mit Überlichtgeschwindigkeit zu senden. Anderseits ist aber einleuchtend, das wir bei strenger Gültigkeit von Newtons Gesetz die Gravitation dazu verwenden könnten, Momentansignale von einem Orte A nach einem entfernten Orte B zu senden; denn die Bewegung einer gravitierenden Masse in A müsste gleichzeitige Änderungen des Gravitationsfeldes in B zur Folge haben - im Widerspruch mit der Relativitätstheorie.

"Die Relativitätstheorie bringt uns aber nicht nur den Zwang Newtons Theorie zu modifizieren; sie schrankt auch zum Glück in weitgehendem Masse die Moglichkeiten hierfür ein. Wäre dies nicht der Fall, so wäre das Bestreben einer Verallgemeinerung von Newtons Theorie ein Hoffnungsloses Unternehmen. Um dies deutlich zu sehen, vesetze man sich nur in folgende analoge Situation: Von den elektromagnetischen Erscheinungen seien nur diejenigen der Elektrostatik experimentell bekannt. Man wisse aber, dass sich die elektrischen Wirkungen nicht mit Uberlichtgeschwindigkeiten ausbreiten können. Wer wäre imstande gewesen, aus diesen Daten die Maxwellsche Theorie der elektromagnetischen Vorgänge zu entwickeln? Unsere Kenntnisse auf dem Gebiete der Gravitation entsprechen aber genau dem eben fingierten Falle; wir kennen nur die Wechselwirkung zwischen ruhenden Massen, und diese wahrscheinlich nur in der ersten Annaherung. Die verwirrende Mannigfaltigkeit der möglichen Verallgemeinerungen wird durch die Relativitätstheorie dadurch eingeschränkt, dass nach letztere in allen Gleichungssystemen die Zeitkoordinate bis aus gewisse Vorzeichenunterschiede in gleicher Weise auftritt wie die drei räumlichen Koordinaten. Diese hier nur ungenau angedeutete grosse formale Ekrenntnis Minkowskis erweist sich beim Aufsuchen von der Relativitätstheorie entsprechenden Gleichungen als Hilfmittel von grösster Bedeutung."

1. A. Einstein, "Zum Gegenwartigen Stande des Gravitationsproblem". Physikalische Zeitschrift, 14 (1913), 1249-50

Here Einstein recognises the incongruity of Newtonian gravitation theory as an action at a distance theory in a climate in which field theories requiring finite propagation times are making great strides forward. But, he notes, this incongruity does not amount to a compulsion to modify Newtonian theory. Further, he sees no decisive empirical reasons for doubting the "strict validity" of the theory. Indeed, in this context, he does not even mention the well known astronomical anomalies. On the contrary, he stresses the success and accuracy of the Newtonian theory in dealing with the motion of astronomical bodies.

The compulsion to modify Newtonian theory only arose with the advent of the special theory of relativity. For, as Einstein notes, that theory forbids the transmission of signals at a speed greater than that of light, in contradiction to the Newtonian theory. Implicit in this is the conclusion that gravitational interactions must now be seen to be mediated by a field.

Finally, Einstein notes that relativity theory powerfully constrains the possible modifications to Newtonian theory. They must all satsify Minkowski's requirement that the time coordinate in the modified theory enter into all equations in the same way as the three spatial coordinates up to differences of sign. In more modern terminology this amounts to the requirement of manifest Lorentz covariance.

3.2 Einstein's Challenge: The Simplest Relativistic Gravitation Theories

In the above passage, Einstein reduces the problem of gravitation in relativity theory to its simplest form through a most suggestive analogy. If we knew only the laws of electrostatics and also the fact that no electrical action can propagate at a speed faster than light, could we recover Maxwell's theory of electromagnetic processes? Up to a change of sign, the laws of electrostatic are identical with those of Newtonian gravitation theory. So in this form the problem of gravitation becomes one of generating a new gravitation theory, which will admit the propagation of no action at a speed faster than light, from Newtonian gravitation theory.

Einstein continues to note that the task of deciding between all the possible generalisations of Newtonian theory is only made feasible by the additional requirement of Lorentz covariance, which is provided by relativity theory. In this section I set out some of the simplest Lorentz covariant generalisations of Newtonian gravitation theory, using a modern four dimensional formalism, in order to examine the prospects of someone who takes up Einstein's challenge and to develop results which will be of use later.

Newtonian gravitation theory can be summarised by the field equation

$$\nabla^2\varphi = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\varphi = 4\pi G\rho \tag{3.1}$$

where φ is the gravitational potential, G the gravitational constant and ρ the mass density, and by the force equation

$$\underline{f} = -m\underline{\nabla}\varphi \tag{3.2}$$

where $\underline{f}$ is the gravitational force acting on a body of mass m and

$$\underline{\nabla} = \underline{i}\frac{\partial}{\partial x} + \underline{j}\frac{\partial}{\partial y} + \underline{k}\frac{\partial}{\partial z}$$

We begin construction of Lorentz covariant generalisations of this theory by replacing the Laplacian operator ∇^2 of field equation (3.1) by the Lorentz covariant d'Alembertian operator $\Box^2 = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}$, which yields a field equation which we can write informally as

$$\Box^2 \varphi = \text{(source density)}$$

It is well known from the properties of the d'Alembertian operator that a field governed by such an equation allows the propagation of disturbances at c, the speed of light in empty space, in accord with Einstein's requirements.

From here the final form of the theory under construction is governed in large measure by the transformation properties of the source density term. If this term is a Lorentz scalar, vector or higher rank tensor, then it follows from the requirement of Lorentz covariance of the field equation that the gravitational potential must be represented by a corresponding scalar, vector or higher rank tensor, φ , φ^{μ}, $\varphi^{\mu\nu}$... respectively.

Mass-energy is the source of the gravitational field. So the first choice for a source density term in the new field equation would be mass-energy density. This quantity transforms as the "time-time" component of the familiar stress-energy tensor $T^{\mu\nu}$ which leads us to select a second rank tensor gravitational potential $\varphi^{\mu\nu}$. and the field equation

$$\Box^2 \varphi^{\mu\nu} = -4\pi G T^{\mu\nu} \qquad (3.3)$$

In a later chapter we shall see this starting point is well chosen, for it can set us on the path to a theory equivalent to Einstein's final general theory of relativity.

It is also possible to choose vector and scalar source density terms. We can choose rest mass, m , to measure the quantity of source of the gravitational field. The coordinate density of this quantity transforms as the "time" component of the four-vector

$$T^{\mu} = \rho U^{\mu}$$

where ρ is the rest mass density and $U^{\mu} = \frac{dx^{\mu}}{d\tau}$ is the four-velocity of the masses concerned with $d\tau^2 = dt^2 - \frac{1}{c^2}(dx^2 + dy^2 + dz^2)$ measuring proper time.

Finally, a natural choice for a scalar source density is rest mass density, ρ . This corresponds to the selection of the term $m\sqrt{1-\frac{v^2}{c^2}}$, where v is the speed of the masses concerned, as the measure of the quantity of source, since rest mass density can be seen to be equal to the density of this quantity, measured in the reference frame in which the masses move at speed v. Another natural candidate for a scalar source density term is the trace of the stress-energy tensor, T^{α}_{α} . This is closely related to ρ , being equal to ρc^2 for a non-interacting distribution of masses. For the present, I will consider the simpler ρ as the scalar source density. The implications of choosing T^{α}_{α} will be discussed in a later chapter.

The choice of the above vector and scalar source densities leads to the following Lorentz covariant theories of gravitation.

<u>The Vector Theory</u>

We write the field equation as

$$\Box^2 \varphi_{\mu} = -4\pi G T_{\mu} \tag{3.4}$$

and construct the simplest Lorentz covariant force equation which generalises equation (3.2):

$$F_\mu = m U^\alpha \varphi_{\alpha,\mu}$$

where F_μ is the four-force acting on a body of rest mass m and four-velocity U^μ. This force equation is unacceptable, for it is inconsistent with the identity of Lorentzian spacetime

$$F_\alpha U^\alpha = m \frac{dU_\alpha}{d\tau} U^\alpha \equiv 0 \tag{3.5}$$

since $m U^\alpha U^\mu \varphi_{\alpha,\mu} \not\equiv 0$

The simplest remedy for this is to introduce a second term into the force equation which now becomes

$$F_\mu = m U^\alpha [\varphi_{\alpha,\mu} - \varphi_{\mu,\alpha}] \tag{3.6}$$

The naturalness of this choice can be verified by noting that it follows from the simple action principle for a mass moving in a vector field

$$\delta \int (mc\sqrt{U_\alpha U^\alpha} + m\varphi_\alpha U^\alpha)\, d\tau = 0$$

The behaviour of this theory is well known since, except for a change of sign, it is formally identical with the basic equations of electrodynamics.

The Scalar Theory

In this case the field equation is

$$\Box^2 \varphi = -4\pi G \rho \tag{3.7a}$$

and the Newtonian force euqation (3.2) naturally generalises to

$$F_\mu = m\, \varphi_{,\mu} \tag{3.8a}$$

where F_μ is the four-force on a body of rest mass m . This force equation suffers the same defect as the corresponding equation in the case of the vector theory; it is inconsistent with the identity

$$F_\alpha U^\alpha = m \frac{dU_\alpha}{d\tau} U^\alpha \equiv 0 \tag{3.5}$$

since $m\,\varphi_{,\mu} U^{\mu} = m \frac{d\varphi}{d\tau} \neq 0$

However, unlike the case of the vector theory, there seems to be no simple modification which can be made to resolve this difficulty while still retaining linearity of the equations in the gravitational potential and its derivatives. However, promising non-linear theories can be constructed by means of a number of devices. I discuss two below.

(1) If m is allowed to vary with φ as $m(\varphi)$, then the identify of Lorentzian spacetime contained in equation (3.5)

$$U^{\alpha} \frac{dU_{\alpha}}{d\tau} \equiv 0$$

no longer entails $F_{\mu} U^{\mu} = 0$

In this case the simple form of the force law (3.8a) can be retained but at the expense of the non-linearity introduced by the variation of m with φ.

Specifically, since $F_{\mu} U^{\mu} = U_{\mu} U^{\mu} \frac{dm}{d\tau} = \frac{d}{d\tau} mc^2$

and from equation (3.8a) $F_{\mu} U^{\mu} = m \frac{d\varphi}{d\tau}$

we are led to $\frac{dm}{d\varphi} = \frac{m}{c^2}$

which integrates to $m(\varphi) = m_0 \exp \frac{\varphi}{c^2}$ (3.9)

where m_0 is a constant peculiar to the mass in question.

This version of the theory follows readily from a simple action principle. The dependence of m on φ enables a simplification of the usual form of such action principles by making possible the fusion of the "particle" and "field-particle" interaction terms into a single term. The action principle is

$$\left.\begin{array}{l} \delta I = 0 \\ \text{where } I = -\frac{1}{8\pi G}\int \varphi^{,\alpha}\varphi_{,\alpha}\, d^4x + \int m(\varphi)\, c\sqrt{U_{\alpha} U^{\alpha}}\, d\tau \end{array}\right\} \quad (3.10a)$$

and $m(\varphi)$ is an undetermined function of φ.

The usual techniques yield a collection of gravitation theories given by the field equation

$$\Box^2 \varphi = -4\pi G c^2 \frac{\partial \rho}{\partial \varphi} \tag{3.7b}$$

and the force equation

$$F_\mu = \frac{d}{d\tau}(m U_\mu) = c^2 \frac{\partial m}{\partial \varphi} \varphi_{,\mu} \tag{3.8b}$$

all of which are consistent with the identity contained in equation (3.5). We are free to stipulate condition (3.9), in which case equations (3.7b) and (3.8b) reduce to the simple forms of equations (3.7a) and (3.8a). This particular theory will be of some importance later for it corresponds to a theory of gravitation constructed by Nordström in 1912.

(2) In the second approach, we retain the independence of m from φ and achieve consistency with the identity contained in equation (3.5) by supplementing the force equation (3.8a) with additional terms. These new versions of the theory can be generated readily be means of the action principle $\delta I = 0$ with

$$I = -\frac{1}{8\pi G}\int \varphi^{,\alpha}\varphi_{,\alpha}\, d^4x + \int m\, f(\varphi)\, c\sqrt{U_\alpha U^\alpha}\, d\tau \tag{3.10b}$$

where $f(\varphi)$ is an undetermined function of φ. This action principle is equivalent to that of (3.10a) with the dependence of m on φ separated in the function f. The usual techniques yield the field equation

$$\Box^2 \varphi = -4\pi G c^2 \rho \frac{df}{d\varphi} \tag{3.7c}$$

and the force equation

$$F_\mu = \frac{d}{d\tau} m U_\mu = c^2 \frac{d \ln f}{d\varphi} m \left[\varphi_{,\mu} - \frac{1}{c^2} U_\mu \varphi_{,\alpha} U^\alpha\right] \tag{3.8a}$$

We are free to set the function f in accord with any further conditions we choose to apply. If we require the force equation to take on the simplest form, we then require

$$\frac{c^2 d \ln f}{d\varphi} = 1$$

which integrates to yield up to an additive constant in φ

$$f(\varphi) = \exp \frac{\varphi}{c^2}$$

With this condition the field equation becomes

$$\Box^2 \varphi = -4\pi G \rho \exp \frac{\varphi}{c^2} \tag{3.7d}$$

and the force equation becomes

$$F_\mu = \frac{d}{d\tau} m U_\mu = m\left[\varphi_{,\mu} - \frac{1}{c^2} U_\mu \varphi_{,\alpha} U^\alpha\right] \tag{3.8d}$$

Substitution of equation (3.9) into (3.7a) and (3.8a) and expansion yields equations identical to (3.7d) and (3.8d), which shows the identity of the two gravitation theories concerned.

3.3 The Gravitation Theories of Poincaré and Minkowski

With the emergence of the special theory of relativity, the unavoidability of modifications to Newtonian gravitation theory became apparent to the general physics community and not just Einstein alone The approach of the general physics community to this problem can be characterised by an examination of the Lorentz covariant theories of gravitation constructed by Einstein's contemporaries, Poincaré and Minkowski, in the first decade of this century.

Poincaré presented his gravitation theory in the ninth section of his celebrated 1905 paper "On the Dynamics of the Electron", a paper in which he formulated the concept of the Lorentz group and invoked the requirement of the covariance of non-electromagnetic laws under this group of transformations.[2] In the first paragraphs of that section Poincaré described how he had been led to seek a Lorentz

---

2. H. Poincaré, "Sur la Dynamique de l'Électron", Comptes Rendus, 140(1905), pp.1504-8; H. Poincaré, "Sur la Dynamique de l'Électron", Rendiconti del Circolo Matematico di Palermo, 21(1906), pp.129-75; for a translation of sections 1-5 and 9 see C.W. Kilmister, Special Theory of Relativity (Oxford: Pergammon, 1970), pp.144-85; for a modernised version see H.M. Schwartz, "Poincaré's Rendiconti Paper on Relativity", American Journal of Physics, 39(1971), pp. 1287-94; 40(1972), pp.862-72, 1282-7; see also A.I. Miller, "A Study of Henri Poincaré's 'Sur la Dynamique de l'Électron'", Archive for History of Exact Sciences, 10(1973), pp.207-328.

covariant theory of gravitation. The "relativity postulate" required, in Poincaré's words, "the impossibility of experimentally demonstrating the absolute motion of the Earth". This postulate had led Lorentz to argue that electromagnetic forces must transform by the Lorentz transformation. Poincaré argued that, if this postulate was to be maintained generally, then a similar result must also obtain for gravitational forces. Otherwise one's state of motion could be determined from the systematic deviations between gravitational and electromagnetic forces in different frames of reference.

Then Poincaré set out exactly how he had constructed his Lorentz covariant gravitation theory. He sought an expression for the gravitational force of attraction between two masses which would satisfy the following conditions:

(1) The relationship describing the propagation of gravitational action (i.e. the retardation time) must be Lorentz covariant.

(2) The force must transform in the same way as electromagnetic forces.

(3) The force law must reduce to the usual Newtonian expression in the case of bodies at rest.

Poincaré recognised that these conditions did not fully specify the gravitation theory. So he imposed the additional conditions:

(4) The law chosen must yield the minimum deviations from Newtonian theory for bodies with small velocities, in accord with the succcess of Newtonian theory in astronomy.

(5) Gravitational action is to propagate forwards in time from a given body.

Poincaré then detailed how he constructed his final law, a more compact statement of which has been given by Harvey. [3] The exact form of the law is not so much of interest to us here. Rather what is important is that it takes the form of an action at a distance law describing the gravitational interaction of two separated masses and also that the law was by no means unique. Poincaré concluded the section by noting how other laws consistent with his conditions could be derived from his original law by a modification of its terms.

Minkowski described and discussed his Lorentz covariant gravitation theory at the end of two of his important papers, his "Basic Equations for Electromagnetic Processes in Moving Bodies" and his Cologne address "Space and Time". [4] In overall conception, Minkowski's theory differed little from that of Poincaré. He sought to find a relativistic generalisation of the Newtonian expression for the gravitational force acting between separated bodies, or, as he would have preferred to describe it, to adapt Newton's law to the four dimensional spacetime world and thus bring it into line with his "world postulate". [5]

---

3. A.L. Harvey, "Brief Review of Lorentz-Covariant Scalar Theories of Gravitation", American Journal of Physics, 33 (1965), pp.449-60.

4. H. Minkowski, "Die Grundgleichungen für die electromagnetischen Vorgänge in bewegten Körpern", Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch - Physikalische Klasse, (1908), pp.53-111;
H. Minkowski, "Raum und Zeit", Physikalische Zeitschrift, 10 (1909), p.104, translated as "Space and Time" in Principle of Relativity (1923; rpt. New York: Dover, 1952), pp.75-91.

5. For Minkowski's discussion of the term "relativity postulate" and his own "world postulate" see Minkowski, "Space and Time", p.83.

We can see the power of Minkowski's introduction of the four dimensional spacetime world here. He sought to construct a gravitational law which would satisfy much the same set of conditions as Poincaré had used, but unlike Poincaré, he did not need to resort to the laborious manipulation of invariant quantities. Rather he could achieve Lorentz covariance quite directly and antomatically by constructing the theory entirely geometrically in four dimensional spacetime. He also ensured correspondence with Newtonian theory by requiring his new law to have the same form as Newton's. Pyenson has given a convenient description and elucidation of the details of the theory. [6]

What is most noticeable about both the theories of Minkowski and Poincaré is their incompleteness. In particular they are both action at a distance theories. This feature is somewhat surprising when it is considered that both Poincaré and Minkowski were extensively involved in the new field approach of electromagnetism and that they both recognised that gravitational action would propagate at a finite speed through space.

We can understand this incompleteness readily once we recognise the motivations underlying their work on gravitation. They were less interested in using relativity theory as a vehicle for developing a new understanding of gravitation. Rather their main concern was to

---

6. L. Pyenson, "The Goettingen Reception of Einstein's General Theory of Relativity", Diss. John Hopkins 1974, pp.201-3; L. Pyenson, "Hermann Minkowski and Einstein's Special Theory of Relativity", Archive for History of Exact Sciences, 17(1977), pp.71-95.

demonstrate that gravitation was not likely to be a source of difficulty for the new relativistic world picture which was then emerging. For it was becoming apparent that the discoveries in electrodynamics were leading to new fundamental results about the nature of space and time which must also apply to all non-electromagnetic phenomena. The most conspicuous of these phenomena was gravitation and if the new world view was to be tenable it must be shown to be consistent with the phenomena of gravitation. Minkowski says as much in introducing his gravitation theory. He writes:

> "I would not like to fail to make it plausible also that a contradiction is not be expected out of the phenomena of gravitation against the acceptance of the relativity postulate." [7]

For the purposes of such a plausibility argument both Poincaré and Minkowski seemed to have found it sufficient to establish that Lorentz covariant generalisations of Newton's law of gravitation can be constructed without problems. This explains Poincaré's apparent satisfaction with his discovery of a range of possible gravitational laws, rather than the one "correct" law. For the more Lorentz covariant laws he could discover, the more plausible became the consistency of gravitation with the postulate of relativity.

Also, in such an exercise, there is no need to develop a detailed theory of the gravitational field. Rather, all that has to be developed of the new gravitation theory are sufficient results to demonstrate consistency with known observations and experiments. As we

---

7. "Ich möchte nicht unterlassen, noch plausibel zu machen, dass nicht von den Erscheinungen der Gravitation her ein Widerspruch gegen die Annahme des Relativitäspostulates zu erwarten ist." Minkowski, "Die Grundgleichungen..." p.109.

have seen, the most rigorous testing ground of gravitation theories lay in the astronomy of the solar system. All that was needed to make predictions in this area was the simple action at a distance laws which they developed.

Consistent with this, both Poincaré and Minkowski turned immediately to the task of working out whether their new laws would lead to significant deviations from the astronomical predictions of Newtonian theory. Poincaré noted his expectation that the deviations introduced would be small, but left the question open for further investigation. He also noted however that his theory introduced deviations which varied with the square of the velocity of the gravitating bodies, whereas Laplace, in his attempt to determine the speed of propagation of gravitation, had introduced deviations varying in direct proportion to this velocity. As we saw earlier, this led Laplace to conclude that the speed of propagation of gravitation must be at least 7×10^6 times that of light, on the basis of known astronomical anomalies. Poincaré could now answer this by noting that his theory predicted deviations 10,000 times smaller than Laplace's for a speed of propagation equal to that of light. [8]

Corresponding to this, Minkowski proceeded to calculate the deviations introduced by his theory from the planetary motions predicted by Newtonian theory. He was able to conclude that they were so small that no decision could be made on this basis against his new theory. [9]

---

8. Poincaré, Comptes Rendus, p.1508 and Rendiconti...,p.175.

9. Minkowski, "Die Grundgleichungen...", pp.110-1 and "Space and Time", p.90.

Another factor may also have led at least Poincaré to limit himself to an action at a distance theory. This was the possibility that all phenomena including gravitation could be reduced to electromagnetism. If this were the case, then the laws of gravitation would automatically be Lorentz covariant and any attempt to produce a comprehensive field theory of gravitation would be superfluous and in danger of being eclipsed by the discovery of the details of this reduction. On this question Poincaré wrote:

> "If we assume the relativity postulate, we find a quantity common to the law of gravitation and the laws of electromagnetism, and this quantity is the velocity of light; and this same quantity appears in every other force, of whatever origin. There can only be two explanations.
>
> Either, everything in the universe is of electromagnetic origin; or, this consituent which appears common to all the phenomena of physcis has no real existence, but arises from our methods of measurement. What are these methods?... In this theory, two lengths are by definition equal if they are traversed by light in the same time.
>
> Perhaps the abandonment of this definition would suffice to overthrow Lorentz's theory..." [10]

This leaves somewhat open the question of whether he believed in an electromagnetic reduction. His alternative, that Lorentz covariance may arise through conventions implicit in our technqiues of measurement, is strongly reminiscent of the appraoch which Einstein was taking to the problem in the same year.

Thus we can conclude that Poincaré and Minkówski approached the question of gravitation principally to show that it was unlikely to cause problems for the new ideas then emerging on space and time.

---

10. Kilmister, p.149.

In a sense we can say that they were dotting the i's and crossing the t's of the new relativistic world picture and that they did not expect any results of great importance to emerge from this. However, at the same time as Minkowski was working on his relativistic gravitation theory, Einstein was to approach the same problem and was to come to a very different conclusion.

3.4 Einstein's 1907 Jahrbuch Paper

In 1907 Einstein wrote an extensive survey article on relativity theory, "On the Principle of Relativity and the Conclusions Drawn From It", in the Jahrbuch der Radioaktivität und Elektronik.[11] In the article Einstein sought to consolidate the position of his new theory. He gave a careful analysis of the foundations of the theory and then continued to indicate how a consistent world picture could emerge from the theory by a careful development of relativistic kinematics and the implications of relativity theory in such areas as electromagnetism, electron theory and thermodynamics.

In the concluding sections Einstein turned to gravitation. In the light of the preceding sections of the paper, it would be natural to expect Einstein to turn to the task of showing that gravitation as well could be expected to meekly submit to the dominion of the new world view, perhaps with arguments similar to those of Poincaré and Minkowski. But this is not what he did.

---

11. A. Einstein, "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen", Jahrbuch der Radioaktivität und Elektronik, 4(1907), pp.411-62; 5(1908), pp.98-9. For a summary of the content and English translation of parts of this paper see H.M. Schwartz, "Einstein's Comprehensive 1907 Essay on Relativity", American Journal of Physics, 45(1977), pp.512-7, 811-7, 899-902.

Einstein began with the speculation that the principle of relativity ought to be extended to encompass accelerated motion. He then postulated the equivalence of gravitation and acceleration and, in an argument which is now well known, suggested that this may enable the extension to be carried out. Out of this equivalence, he constructed a theory of gravitation which dealt in particular with the influence of gravitation on the speed of clocks and on electromagnetic processes. He concluded that a gravitational field would affect the speed at which clocks ran and change the speed and direction of light propagation.

Knowing what Einstein will ultimately make of these ideas, the modern reader can easily underestimate their audacity. At this time relativity theory was still struggling to find its feet and this was especially the case with Einstein's particular version of it. Einstein was still working in the Bern patents office and his 1905 paper was not yet known to many physicists let alone acclaimed, as it was later to be.

It is against this background that Einstein introduced results which directly contradict the basic assumption of his 1905 theory of relativity.[12] The variability in the rate of clocks contradicted Einstein's 1905 assumption of the homogeneity of time, for, in 1905, Einstein had reduced time to little more than the set of clock readings. Yet more serious was Einstein's conclusion that light propagation could vary in both speed and direction. For the controversial requirement of the constancy of the velocity of light

12. As in A. Einstein, "On the Electrodynamics of Moving Bodies" in <u>The Principle of Relativity</u> (New York: Dover, 1952), pp.35-65.

was fundamental to the theory and, at least prior to 1907, quite unnegotiable. Indeed it was on the basis of this uncompromising constancy that light could become the basic measuring stick of the theory and guide Einstein through the perplexing relativities of space and time.

Einstein could offer surprisingly little to justify this bold departure. There was the possibility of an extension of the principle of relativity, but, in the 1907 paper, it remained exactly that, a possibility. In the original article he could present no sustained vision of how this extension was to come about. As we shall see, in spite of later elaborations to this basic theory of gravitation, it was not until 1913 that Einstein was able to publish any major advance on his first attempts to realise this extension.

In view of this, it is not at all surprising that Einstein should become involved in a vitriolic dispute with Abraham in 1912 over whether Einstein's developing gravitation theory did in fact amount to the rejection of his earlier special theory of relativity, or an extension of it, as Einstein claimed.[13]

---

13. M. Abraham, "Das Gravitationsfeld", Physikalische Zeitschrift, 13(1912), pp.793-7; M. Abraham, "Relativität und Gravitation. Erwiderung auf eine Bermerkung des Hrn. A. Einstein", Annalen der Physik,38(1912) pp.1056-8; M. Abraham, "Nochmals Gravitation und Relativität. Bermerkungen zu A. Einsteins Erwiderung", Annalen der Physik,39(1912), pp.444-8; A. Einstein, "Relativität und Gravitation: Erwiderung auf eine Bemerkung von M. Abraham", Annalen der Physik,38(1912), pp.1059-64; A. Einstein "Bemerkung zu M. Abrahams vorangehender Auseinandersetzung 'Nochmals Relativität und Gravitation'", Annalen der Physik,39(1912), p.704.

The remainder of the chapter will be devoted to the question of how Einstein came to the remarkable conclusion that gravitation could not be accommodated within the framework of the new ideas on space and time then emerging - a conclusion which directly contradicted the viewpoints of some of Einstein's most distinguished contemporaries.

3.5 The Starting Point

In his 1907 Jahrbuch...article, Einstein speculated that his new gravitation theory would enable an extension of the relativity postulate to accelerated motion. Since we know that his gravitation theory would ultimately blossom into his general theory of relativity, it is tempting to conclude that the starting point historically for Einstein's work on gravitation was the hope that it may enable just such a theory to emerge.

The testing of this inference is made difficult by the fact that we have virtually no primary source material on how Einstein came to his 1907 gravitation theory. In particular, the 1907 article was his first publication on the problem of gravitation in relativity theory. Perhaps his first extant mention of this problem is in a letter to Conrad Habicht dated 24th December; there Einstein wrote:

> "In October and November I was extremely busy on a partly revisional, partly new work on the relativity principle. I send you the results. Now I am busy on a relativistic theory of the gravitation law with which I hope to account for the still unexplained secular changes of the perihelion movement of mercury. So far I have not managed to succeed." 14

14. Quoted in C. Seelig, Albert Einstein. A Documentary Biography, trans. M. Savill (London: Staples, 1956), p.76.

The "partly revisional, partly new work" referred to is clearly the 1907 Jahrbuch... article. The comments which follow this lead to the speculation that Einstein may have decided to construct a relativisitic gravitation theory in order to try to account for the anomalous motion of Mercury.

However, in a wider context, there is little further evidence to support this idea. Nowhere else does Einstein discuss the anomalous motion of Mercury in relation to his original work on gravitation. It is not mentioned in the 1907 Jahrbuch...article and is only dealt with seriously in his publications in 1915, when to his jubilation he found that the nearly completed general theory of relativity could account exactly for the known anomaly.[15] Moreover we have already seen in Einstein's 1913 comments quoted earlier that he believed that there was no reason to doubt the strict "Validity" of Newtonian gravitation theory as far as experience was concerned.

This, however, casts little light on the starting point of Einstein's work in this area. Did he turn to the problem of gravitation because he saw in it a means of solving the problem of extending the principle of relativity? Or did he first work on the problem of gravitation and then realise that it brought the possibility of extending the relativity principle? If we sift through the texts of Einstein's many lectures on and popularisations of relativity theory we find a sufficient number of overtly historical statements to enable us to decide in favour of the latter of these two cases.

---

15. A. Einstein, "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie", Preussische Akademie der Wissenschaften, Sitzungsberichte, 1915 Part 2, pp.831-9.

In an account of the origins of relativity theory written by Einstein in about 1919, he wrote:

> "When, in the year 1907, I was working on a summary essay concerning the special theory of relativity for the Jahrbuch für Radioaktivität und Elektronik [sic], I had to try to modify Newton's theory of gravitation in such a way that it would fit into the theory [of relativity]. Attempts in this direction showed the possibility of carrying out this enterprise, but they did not satisfy me because they had to be supported by hypotheses without physical basis. At that point there came to me the happiest thought of my life..."

This was followed by a discussion of the interchangeability of gravitation and acceleration which led up to the conclusion:

> "The fact of experience concerning the independence of acceleration in free fall with respect to the material is therefore a mighty argument that the postulate of relativity is to be extended to co-ordinate systems that move non-uniformly relative to one another..." [16]

In a short paper on the historical origins of his general theory called "Notes on the Origin of the General Theory of Relativity", he wrote:

> "From the purely kinematical point of view there was no doubt about the relativity of all motion whatever; physically speaking, the inertial system seemed to occupy a privileged position, which made the use of coordinate systems moving in other ways appear artificial.
>
> "I was of course aquainted with Mach's view, according to which it appeared conceivable that what inertial resistance counteracts is not acceleration as such but acceleration with respect to the masses of the other bodies existing in the world. There was something fascinating about this idea to me, but it provided no workable basis for a new theory.

---

16. A. Einstein, "Fundamental Ideas and Methods of Relativity Theory, Presented in their Development", Unpublished Manuscript at Einstein Archives, Princeton, circa 1919. The translation of the title of the paper and of the passage quoted from the original German is taken from G. Holton, "Finding Favour with the Angel of the Lord. Notes Towards the Psychobiographical Study of Scientific Genius" in Y. Elkana, ed., The Interaction between Science and Philosphy (Humanities Press, 1975), pp.369-71.

> "I first came a step nearer to the solution of the problem when I attempted to deal with the law of gravity within the framework of the special theory of relativity. Like most writers at the time, I tried to frame a field-law for gravitation, since it was no longer possible, at least in any natural way, to introduce direct action at a distance owing to the abolition of the notion of absolute simultaneity."

He continued to describe his attempt to produce a gravitation theory within special relativity, how such a theory contradicted "the principle of the equality of inertial and gravitational mass" and how he "abandoned as inadequate the attempt to treat the problem of gravitation, in the manner outlined above, within the framework of the special theory of relativity."

He concluded:

> "If this principle held good for any events whatever (the 'principle of equivalence'), this was an indication that the principle of relativity needed to be extended to co-ordinate systems in non-uniform motion with respect to each other, if we were to reach a natural theory of gravitational fields. Such reflections kept me busy from 1908 to 1911, and I attempted to draw special conclusions from them, of which I do not propose to speak here. For the moment the one important thing was the discovery that a reasonable theory of gravitation could only be hoped for from an extension of the principle of relativity." [17]

Finally, we find a very similar account in Einstein's "last lecture", a lecture given at Princeton University on the 14th April, 1954. After discussing the unacceptability of inertial systems as absolutes in physics, he continued to describe how he came to see gravitation as the means of removing inertial systems from this preferred status. I quote from the notes taken by J. A. Wheeler:

> "This is not the direct way I found the theory of gravitation. The real way is a very strange story.
>
> "I had to write a paper about the content of special relativity. Then I came to the question of how to handle gravity. The object falls with a different acceleration if it is moving than if it is not moving...Thus a gas

---

17. A. Einstein, "Notes on the Origin of the General Theory of Relativity" in Ideas and Opinions (London: Souvenir, 1973),pp.285-90. Einstein's emphasis.

> falls with another acceleration if heated than if not heated. I felt this is not true. Came out that acceleration is independent of quality of matter: pendulum experiments.
>
> Change coordinate systems? Then change acceleration. Then I came to a real understanding of the equivalence of gravitational and inertial mass. No inertial system can be preferred. That was not so clear to me at the time..."[18]

These extracts all point to the same account of the starting point of Einstein's 1907 speculations on gravitation and relativity. After 1905, Einstein felt some dissatisfaction with the fact that the special theory of relativity had asserted the relativity of motion only as far as inertial motion was concerned. It is not clear to what extent he was explicitly aware of this dissatisfaction. Certainly he could propose no new theory which would resolve it. Then, as a part of his 1907 Jahrbuch...survey article, he tried to bring the phenomenon of gravitation into the relativistic world picture, presumably with much the same motives as Poincaré and Minkowski. He found, however, that he could construct no acceptable gravitation theory within special relativity. It appeared that gravitation could not be accounted for within the framework of special relativity. Focussing on the cause of this failure, he introduced his belief that the relativity principle must be extended in order to eliminate the absolute status of acceleration and inertial systems and turned to developing a gravitation theory which would enable this extension to be realised.

So far, I have not discussed the crucial step in this argument, Enstein's conclusion that no acceptable gravitation theory could be formulated within special relativity. In the following sections, I turn to discuss this question.

---

18. J.A. Wheeler, "Einstein's Last Lecture" in G.E. Tauber, ed., Albert Einstein's Theory of General Relativity (New York: Crown, 1979) pp.188-90. Einstein's emphasis.

3.6 Einstein's Rejection of Special Relativistic Gravitation Theories

3.6.1 Einstein's 1907 Special Relativistic Theory of Gravitation

In a number of places, Einstein has described how he tried to erect a special relativistic gravitation theory whilst preparing the Jahrbuch...article and why he felt that such a theory was unnacceptable. In the short historical paper, "Notes on the Origin of the General Theory of Relativity", he wrote in continuation of the extract given in the last section:

> "...I attempted to deal with the law of gravitation within the framework of the special theory of relativity. Like most writers at the time, I tried to frame a field-law for gravitation, since it was no longer possible, at least in any natural way, to introduce direct action at a distance owing to the abolition of the notion of absolute simultaneity.
>
> The simplest thing was, of course, to retain the Laplacian scalar potential of gravity, and to complete the equation of Poisson in an obvious way by a term differentiated with respect to time in such a way that the special theory of relativity was satisfied. The law of motion of the mass point in a gravitational field had also to be adapted to the special theory of relativity. The path was not so unmistakably marked out here, since the inert mass of a body might depend on the gravitational potential. In fact, this was to be expected on account of the principle of the inertia of energy.
>
> These investigations, however, led to a result which raised my strong suspicions. According to classical mechanics, the vertical acceleration of body in the vertical gravitation field is independent of the horizontal component of its velocity. Hence in such a gravitational field the vertical acceleration of a mechanical system or of its center of gravity works out independently of its internal kinetic energy. But in the theory I advanced, the acceleration of a falling body was not independent of its horizontal velocity or the internal energy of system.
>
> This did not fit in with the old experimental fact that all bodies have the same acceleration in a gravitational field. This law, which may also be

formulated as the law of the equality of inertial and gravitational mass, was now brought home to me in all its significance. I was in the highest degree amazed at its existence and guessed that in it must lie the key to a deeper understanding of inertia and gravitation..." 19

In a similar passage in his Autobiographical Notes, Einstein wrote:

"That the special theory of relativity is only the first step of a necessary development became completely clear to me only in my efforts to represent gravitation in the framework of this theory. In classical mechanics, interpreted in terms of the field, the potential of gravitation appears as a scalar field (the simplest theoretical possibility of a field with a single component). Such a scalar theory of gravitation can easily be made invariant under the group of Lorentz transformations. The following program appears natural, therefore: the total physical field consists of a scalar field (gravitation) and a vector field (electromagnetic field); later insights may eventually make necessary the introduction of still more complicated types of fields; but to begin with one did not need to bother about this.

The possibility of realization of this program was, however, in doubt from the very first, because the theory had to combine the followings things:

(1) From the general considerations of special relativity theory it was clear that the inertial mass of a physical system increases with the total energy (therefore, e.g., with the kinetic energy).

(2) From very accurate experiments (especially from the torsion balance experiments of Eötvös) it was empirically known with very high accuracy that the gravitational mass of a body is exactly equal to its inertial mass.

It followed from (1) and (2) that the weight of a system depends in a precisely known manner on its total energy. If the theory did not accomplish this or could not do it naturally, it was to be rejected. The condition is most naturally expressed as follows: The acceleration of a system falling freely in a given gravitational field is independent of the nature of the falling system (especially therefore also of its energy content).

---

19. A. Einstein, "Notes on the Origin of the General Theory of Relativity," pp.286-7. Einstein's emphasis.

> It turned out that, within the framework of the program sketched, this simple state of affairs could not at all, or at any rate not in any natural fashion, be represented in a satisfactory way. This convinced me that within the structure of the special theory of relativity there is no niche for a satisfactory theory of gravitation." [20]

These passages raise a number of questions. What was the special relativistic gravitation theory which Einstein erected in 1907? In particular, what was its field equation and its force equation and how did it incorporate the dependence of inertial mass on gravitational potential? Finally, can we recover the prediction which Einstein felt rendered the theory unnacceptable and, if so, how are we to understand this result? I deal with these questions in order below.

The Field Equation

It can be seen that the theory of gravitation referred to by Einstein in these extracts is closely related to the simple scalar theories of gravitation outlined earlier in the chapter. In those theories, construction of the field equation was commenced by generalising the field equation of Newtonian theory

$$\nabla^2\varphi = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\varphi = 4\pi G\rho \qquad (3.1)$$

as the natural Lorentz covariant field equation

$$\Box^2\varphi = \left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}\right)\varphi = -4\pi G\rho \qquad (3.7a)$$

This is clearly the modification which Einstein was describing when he wrote in the first extract of this section:

> "The simplest thing was, of course, to retain the Laplacian scalar potential of gravity, and to complete the equation of Poisson [equation (3.1) above] in an obvious way by a term differentiated with respect to time..."

---

20. A. Einstein, Autobiographical Notes, trans. P.A. Schilpp (La Salle and Chicago, Ill.: Open Court. 1979), pp.59-61. Einstein's emphasis

The Force Law

Einstein continues to note that the path to the law of motion of the mass point "...was not so unmistakably marked out...". We have already seen some of these difficulties earlier. In a Lorentz covariant formalism, the most natural force law is

$$F_\mu = m\,\varphi_{,\mu} \qquad (3.8a)$$

in which m is assumed to be independent of φ . This contradicted the identity

$$F_\alpha U^\alpha = m \frac{dU_\alpha}{d\tau} U^\alpha \equiv 0 \qquad (3.5)$$

and we saw that this contradiction could be resolved by

(1) allowing m to vary with φ , which enables force law (3.8a) to be retained provided $m(\varphi) = m_0 \exp \frac{\varphi}{c^2}$ (3.9)

or

(2) adding extra terms to the force law (3.8a), for example, of the form

$$F_\mu = m\left[\varphi_{,\mu} - \frac{1}{c^2} U^\alpha \varphi_{,\alpha} U_\mu\right] \qquad (3.8d)$$

If Einstein did begin with a force law equivalent to equation (3.8a), then his comments on the dependence of the inertial mass of a body on the gravitational potential suggest that he followed option (1).

However, it is by no means clear that Einstein did use this force law. The "naturalness" of this law depends very much on the formalism within which it is written and, at this time, the relevant four-vector formalism was only just emerging in the work of Minkowski Einstein did not use this formalism in his 1907 Jahrbuch... paper; rather he used essentially the same formal techniques as in his 1905 relativity paper. So presumably he used these same techniques in his work on the problem of gravitation in special relativity.

The techniques which Einstein used in the 1907 Jahrbuch... paper to construct the Lorentz force law of electrodynamics (Equation (4.7a) below) can be applied almost unchanged in the gravitational case and can lead to force laws equivalent to those discussed above. I turn now to the task of reconstructing this. It seems likely that Einstein would have carried out a similar analysis, although, of course, no guarantee can be given. In any case, a return to the formalism he would have used can only serve to bring us closer to his thinking. Also, the returning of the force laws developed earlier from a different technique, suggests that they have some intrinsic significance, independent of the formalism used, and therefore that Einstein would most likely have arrived at them by whatever other method he adopted.

In the following, I show Einstein's construction of the Lorentz force law as given in the 1907 Jahrbuch... article on the left of the page and the analogous steps in the construction of a gravitational force law on the right. Page numbers refer to the 1907 Jahrbuch... article. I temporarily revert to the use of Einstein's notation in both cases.

Electrodynamic case

(X,Y,Z) is the electric field strength vector
(L,M,N) is the magnetic field strength vector

Gravitational case

φ is the gravitational potential
$(Xg,Yg,Zg) = \left(-\frac{\partial\varphi}{\partial x}, -\frac{\partial\varphi}{\partial y}, -\frac{\partial\varphi}{\partial z}\right)$ (3.11)
is the gravitational field strength vector.

Thus the field equation (3.7a) can be written as

$$\Box^2\varphi = \frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} + \frac{\partial X_g}{\partial x} + \frac{\partial Y_g}{\partial y} + \frac{\partial Z_g}{\partial z}$$
$$= -4\pi G\rho$$

Consider a coordinate system S'(x',y',z') with primed quantities and whose origin moves with velocity v in the x-direction of a coordinate system S(x,y,z) with unprimed quantities.

From the requirement that Maxwell's equations be covariant with respect to a Lorentz transformation, it can be concluded that field vectors transform according to

$$\left.\begin{aligned} X' &= X \\ Y' &= \beta\left(Y - \tfrac{v}{c} N\right) \\ Z' &= \beta\left(Z + \tfrac{v}{c} M\right) \end{aligned}\right\} \quad (3.12a)$$

$$\left.\begin{aligned} L' &= L \\ M' &= \beta\left(M + \tfrac{v}{c} Z\right) \\ N' &= \beta\left(N - \tfrac{v}{c} Z\right) \end{aligned}\right\} \quad (3.13)$$

(pp.427-8)

From the definition of the gravitational field strength vector, it can be concluded that this vector transforms according to

$$\left.\begin{aligned} X'_g &= \beta\left(X_g - \tfrac{v}{c^2} \tfrac{\partial \varphi}{\partial t}\right) \\ Y'_g &= Y_g \\ Z'_g &= Z_g \end{aligned}\right\} \quad (3.12b)$$

It can be readily confirmed that this transformation law is consistent with the Lorentz covariance of the field equation (3.7a)

with $\beta = \dfrac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$

A body with rest mass μ and space coordinates (x_o, y_o, z_o) is instantaneously at rest in S'. Its equations of motion are:

$$\mu \frac{d^2 x'_o}{dt^2} = \varepsilon X' \qquad \mu \frac{d^2 y'_o}{dt^2} = \varepsilon Y' \qquad \mu \frac{d^2 z'_o}{dt^2} = \varepsilon Z'$$

for a body with charge ε (pp.431-2)

$$\mu \frac{d^2 x'_o}{dt^2} = \mu X'_g \qquad \mu \frac{d^2 y'_o}{dt^2} = \mu Y'_g \qquad \mu \frac{d^2 z'_o}{dt^2} = \mu Z'_g$$

These equations are transformed to S by substitution from the transformation forumulae for the quantities in each equation and noting that for the particular case at hand

$$\frac{d^2 x'_o}{dt^2} = \beta^3 \ddot{x} \qquad \frac{d^2 y'_o}{dt^2} = \beta^2 \ddot{y} \qquad \frac{d^2 z'_o}{dt^2} = \beta^2 \ddot{z} \qquad \text{where} \quad \dot{x} = \frac{dx}{dt} \text{ etc.}$$

We find

$$\mu\beta^3 \ddot{x}_o = \varepsilon X$$

$$\mu\beta \ddot{y}_o = \varepsilon\left(Y - \tfrac{v}{c} N\right)$$

$$\mu\beta \ddot{z}_o = \varepsilon\left(Z + \tfrac{v}{c} N\right) \quad \text{(p.432)}$$

$$\mu\beta^3 \ddot{x}_o = \mu\beta\left(X_g - \tfrac{v}{c^2} \tfrac{\partial \varphi}{\partial t}\right)$$

$$\mu\beta \ddot{y}_o = \mu\sqrt{1 - \tfrac{v^2}{c^2}}\, Y_g$$

$$\mu\beta \ddot{z}_o = \mu\sqrt{1 - \tfrac{v^2}{c^2}}\, Z_g$$

These equations can be worked up to a form symmetrical in the x,y and z coordinates if we make use of the following definition of force (K_x, K_y, K_z)

$$K_x = \frac{d}{dt}\left\{\frac{\mu \dot{x}}{\sqrt{1-q^2/c^2}}\right\} \qquad K_y = \frac{d}{dt}\left\{\frac{\mu \dot{y}}{\sqrt{1-q^2/c^2}}\right\} \qquad K_z = \frac{d}{dt}\left\{\frac{\mu \dot{z}}{\sqrt{1-q^2/c^2}}\right\} \tag{3.14}$$

where $q = \sqrt{\dot{x}_o^2 + \dot{y}_o^2 + \dot{z}_o^2}$

is the velocity of the body.

It is interesting to note that Einstein did not use this definition of force in his 1907 relativity paper. Its use in the 1908 Jahrbuch... article greatly simplified his account of the mechanics of a point mass. It no longer required him to distinguish transverse and longitudinal mass and made possible the writing of the laws of conservation of energy and momentum in a very simple form. As Einstein acknowledged, this definition had been introduced first by Planck.[21]

The equations of motion in S are rewritten, dropping the subscript o from the position coordinates and making use of the above definition of force and the equalities.

$$\frac{d}{dt}\left(\frac{\dot{x}}{\sqrt{1-q^2/c^2}}\right) = \beta^3 \ddot{x} \qquad \frac{d}{dt}\left(\frac{\dot{y}}{\sqrt{1-q^2/c^2}}\right) = \beta \ddot{y} \qquad \frac{d}{dt}\left(\frac{\dot{z}}{\sqrt{1-q^2/c^2}}\right) = \beta \ddot{z}$$

(It can be confirmed that these equations are identities for the case of $\dot{x} = q$, $\dot{y} = 0$ and $\dot{z} = 0$.)

The equations which result from this are

$$\left.\begin{aligned} K_x &= \varepsilon X \\ K_y &= \varepsilon\left\{Y - \frac{\dot{x}}{c} N\right\} \\ K_z &= \varepsilon\left\{Z + \frac{\dot{x}}{c} M\right\}\end{aligned}\right\} \tag{3.15a}$$

Inserting corresponding terms in $\dot{y}$ and $\dot{z}$, in order to recover the force law for the general case when $\dot{y}$ and $\dot{z}$ are not zero, we construct the Lorentz force law:

$$\left.\begin{aligned} K_x &= \mu\beta\left(X_g - \frac{\dot{x}}{c^2}\frac{\partial \Phi}{\partial t}\right) \\ K_y &= \mu\sqrt{1-q^2/c^2}\, Y_g \\ K_z &= \mu\sqrt{1-q^2/c^2}\, Z_g\end{aligned}\right\} \tag{3.15b}$$

There appears to be no simple analogous method for extending these equations to a form which is symmetrical in the x,y and z coordinates.

These equations readily revert to such a form, however, in the case of one special condition: that the gravitational field is static when viewed from the body's rest frame.

---

21. Einstein, "Über das Relativitätsprinzip...", p.414; M. Planck, Verh. Deutsch. Phys. Ges., 8 (1906), p.136.

This entails $\frac{d\varphi}{d\tau} = 0$ (3.16)

with τ = proper time.

Since S' is the body's rest frame

$$\frac{\partial\varphi}{\partial t'} = 0$$

which transforms to

$$\beta\left(\frac{\partial\varphi}{\partial t} + v\frac{\partial\varphi}{\partial x}\right) = 0$$

in S. With this condition, the force law in S becomes

$$\left.\begin{aligned} K_x &= \varepsilon\left\{X + \frac{\dot{y}}{c}N - \frac{\dot{z}}{c}M\right\} \\ K_y &= \varepsilon\left\{Y + \frac{\dot{z}}{c}L - \frac{\dot{x}}{c}N\right\} \\ K_z &= \varepsilon\left\{Z + \frac{\dot{x}}{c}M - \frac{\dot{y}}{c}L\right\} \end{aligned}\right\} \quad (3.17a)$$

(pp.432-3)

$$\left.\begin{aligned} K_x &= \mu\sqrt{1 - q^2/c^2}\, X_g \\ K_y &= \mu\sqrt{1 - q^2/c^2}\, Y_g \\ K_z &= \mu\sqrt{1 - q^2/c^2}\, Z_g \end{aligned}\right\} \quad (3.17b)$$

The resulting gravitational force law (3.17b) is the same as the spatial part of the four-vector force law

$$F_\mu = m\,\varphi_{,\mu} \qquad (3.8a)$$

This follows from the relationship between the spatial components of the four-force and those of the force vector defined in equation (3.14):

$$F_1 = \frac{-K_x}{\sqrt{1-q^2/c^2}} \qquad F_2 = \frac{-K_y}{\sqrt{1-q^2/c^2}} \qquad F_3 = \frac{-K_z}{\sqrt{1-q^2/c^2}}$$

It is interesting that this result could be obtained only in the special case when

$$\frac{d\varphi}{d\tau} = 0$$

for it is only in this special case that force law (3.8a) satisfies the conditions

$$F_\mu U^\mu = 0 \qquad (3.5)$$

since $\quad F_\mu U^\mu = m\varphi_{,\mu}U^\mu = m\frac{d\varphi}{d\tau}$

This suggests that the problem faced in recovering a law symmetrical in the x,y and z coordinates from equation (3.15b) is essentially the same as the one involved in the modification of force law (3.8a) to make it consistent with the condition (3.5). I show below that the two approaches to the solution of this latter problem do in fact recover symmetric force laws from equation (3.15b). The reason for this can be seen once it is recognised that the identity (3.5), when written in the relevant three dimensional formalism, is merely the well known energy law

$$\frac{d}{dt}\left(\frac{\mu c^2}{\sqrt{1-v^2/c^2}}\right) = K_x\dot{x} + K_y\dot{y} + K_z\dot{z} \tag{3.5a}$$

which, in less formal terms, amounts to the requirement

(Energy) = (Force). (Distance)

In this case, equation (3.5a) can be looked upon as a severe constraint on allowable definitions of force, requiring in particular that force be defined in a way that is symmetrical with respect to the space coordinates.

Further, the achieving of consistency between the new force law and the energy law (3.5a) follows naturally as the next step of the method embodied in Einstein's 1907 Jahrbuch... treatment of electrodynamics. Immediately following the establishment of the Lorentz force law (3.17a), Einstein turned to the task of showing that the theory was consistent with the principles of conservation of energy and momentum. Part of this involved the displaying of the energy law (3.5a) in its integrated form.[22] Thus any inconsistency between this law and the force law would automatically surface and require attention.

---

22. Einstein, "Über das Relativitätsprinzip...", p.434 - Equation 14.

require attention.

We have seen that this consistency can be achieved in two ways. The second involved the retention of the independence of rest mass from the gravitational potential and the supplementing of the force law (3.8a)/(3.17b) with extra terms. One of the simplest of these modifications was given in equation (3.8d). In the relevant formalism, its spatial part becomes

$$\left.\begin{aligned} K_x &= -\mu\left[\sqrt{1-q^2/c^2}\,\frac{\partial\varphi}{\partial x} + \frac{\beta\dot{x}}{c^2}\left(\frac{\partial\varphi}{\partial t} + \dot{x}\frac{\partial\varphi}{\partial x} + \dot{y}\frac{\partial\varphi}{\partial y} + \dot{z}\frac{\partial\varphi}{\partial z}\right)\right] \\ K_y &= -\mu\left[\sqrt{1-q^2/c^2}\,\frac{\partial\varphi}{\partial y} + \frac{\beta\dot{y}}{c^2}\left(\frac{\partial\varphi}{\partial t} + \dot{x}\frac{\partial\varphi}{\partial x} + \dot{y}\frac{\partial\varphi}{\partial y} + \dot{z}\frac{\partial\varphi}{\partial z}\right)\right] \\ K_z &= -\mu\left[\sqrt{1-q^2/c^2}\,\frac{\partial\varphi}{\partial z} + \frac{\beta\dot{z}}{c^2}\left(\frac{\partial\varphi}{\partial t} + \dot{x}\frac{\partial\varphi}{\partial x} + \dot{y}\frac{\partial\varphi}{\partial y} + \dot{z}\frac{\partial\varphi}{\partial z}\right)\right] \end{aligned}\right\} \quad (3.8e)$$

This new force law is satisfactory in so far as it reduces to equation (3.15b) for the special case of $\dot{x} = q$, $\dot{y} = 0$ and $\dot{z} = 0$. However, the law is more complex than would be expected and is not directly suggested by equation (3.15b). If, however, we were to follow the first option, that of allowing rest mass to vary with gravitational potential, then it would be possible to retain the simpler force law (3.17b), a law which is closer in form to the original Newtonian force law.

If rest mass μ is allowed to vary with gravitational potential φ, we now have

$$K_x = \frac{d}{dt}\left(\frac{\mu\dot{x}}{\sqrt{1-q^2/c^2}}\right) = \mu\beta^3\ddot{x} + \beta\dot{x}\frac{d\mu}{dt}$$

whilst retaining $K_y = \beta\mu\ddot{y}$ and $K_z = \beta\mu\ddot{z}$

for the special case of $\dot{y} = 0$ and $\dot{z} = 0$. This leads to the replacement of equation (3.15b) by

$$\left.\begin{aligned} K_x &= \mu\beta\left(X_g - \frac{\dot{x}}{c^2}\frac{\partial\varphi}{\partial t}\right) + \beta\dot{x}\frac{d\mu}{dt} \\ K_y &= \mu\sqrt{1-q^2/c^2}\,Y_g \\ K_z &= \mu\sqrt{1-q^2/c^2}\,Z_g \end{aligned}\right\} \quad (3.15c)$$

This reduces to the symmetrical form of equation (3.17b) if and only if $\frac{d\mu}{d\varphi} = \frac{\mu}{c^2}$ which integrates to $\mu(\varphi) = \mu(0) \exp\frac{\varphi}{c^2}$
Thus, provided μ varies with φ in this way, the force law (3.17b) can be retained for the general case. This corresponds exactly to the retention of force law (3.8a) by means of condition (3.9) in the four-vector presentation of this theory.

This version of the theory is the one that follows most naturally from the methods of the 1907 Jahrbuch...paper, whilst retaining consistency with Einstein's other comments. In particular, it entails the dependence of inertial mass $\beta\mu$ and, correspondingly, energy $\beta\mu c^2$ on the gravitational potential as a result of relation (3.9). This is consistent with Einstein's

> "...inert mass of a body might depend on the gravitational potential. In fact, this was to be expected on account of the principle of the inertia of energy." [23]

It is also the simplest theory which contains the dependence of inertial mass on gravitational potential. No such dependence arises in the case of theories, which are based on the addition of extra terms to the basic force law (3.17b) such as in the force law (3.8e).

From the material available, it is only possible to speculate on the form of the theory that Einstein constructed and on the extent to which he followed through the arguments presented above. Fortunately, as we shall see, this does not create any great difficulty for the task of understanding why Einstain felt his theory - and any special relativistic theory - to be unnacceptable.

---

23. See note 19

The Rejection of the Theory

In the Einstein quotes given at the beginning of this section, section (3.6.1), Einstein described why he felt that the theory he had constructed was unacceptable. The rejection focussed on the fact fact that the theory predicted that the acceleration of a falling body would not be independent of its velocity or its energy content. We can show how this follows from the theories sketched out above if we consider the special case of a static gravitational field, directed only in the x-direction, acting on a body which instantaneously has velocity in the y-direction alone.

The original force law (3.17b) was derived for the case of a static field and was independent of the assumptions used later to generate the general force laws. Unfortunately, in this form it cannot be applied directly to the case considered here, for it was derived for a field which was static in the moving body's rest frame. But if we apply the general force laws to this case - that is force law (3.17b) as a general law with rest mass varying according to condition (3.9) or force law (3.8e) - then in both cases we arrive at the equations

$$K_x = -\mu\sqrt{1 - q^2/c^2}\,\frac{\partial\varphi}{\partial x} \qquad K_y = K_z = 0 \tag{3.18}$$

where $q = \dot{y}$. I suggest below that any force law developed within the constraints used above will yield such a result.

From equation (3.18), it follows directly that the initial "downward" acceleration of the body in the x-direction will depend on the "sideways" velocity of the body in the y-direction. For, as this sideways velocity of the body increases, the gravitational force acting on the body decreases. Yet, at the same time, the inertia of the body increases. Thus a body with faster sideways velocity will fall slower.

This can also be interpreted as a dependence of the rate of fall on the energy of the system. The downward acceleration of the body decreases with an increase in the kinetic energy of the body, due to its sideways motion. This can be extended to a case that Einstein considered elsewhere.[24] An increase in the internal energy of a gas due to heating, which amounts to an increase in the kinetic energy of the gas molecules, leads to a decrease in the downward acceleration of the gas.

Einstein felt that these predictions contradicted the observation, recognised since the time of Galileo, that all bodies fall with the same acceleration in the earth's gravitational field.

Einstein also described the failure of the theory in other terms. From the observation mentioned above, he was led to expect that the weight or gravitational mass of the body should vary in proportion to its total energy. This, he concluded, was not the case in the theory. Approaching such an objection is a little awkward for the term "gravitational mass" is a theoretical term, which is well defined in Newtonian gravitation theory but is undefined by Einstein in this context.

However, the nature and persistence of the appearance of the term $\mu\sqrt{1-q^2/c^2}=m\sqrt{1-v^2/c^2}$ in such force laws as (3.17b) and (3.18) enable

---

24. Wheeler, p.188.

us to select it unambiguously as the term representing gravitational mass, the measure of gravitational source. It is clear from the discussion earlier in the chapter that the recovery of this term stems directly from the requirement that the gravitational potential be represented by a scalar quantity in a Lorentz covariant theory. For this reason we would expect to find this quantity measuring gravitational mass in any scalar Lorentz covariant theory and, in particular, to recover equation (3.18) from the majority of admissible force laws.

Further, we have seen how the selection of the term representing gravitational source, in conjunction with the requirement of Lorentz covariance, governs the nature of the resulting gravitation theory. The selection of rest mass, m , leads to a vector theory and the selection of mass, $\frac{m}{\sqrt{1-v^2/c^2}}$, leads to a second rank tensor theory. Thus we would only expect second rank tensor theories to be able to satisfy Einstein's requirement that gravitational mass vary in direct proportion to total energy.

So in retrospect, Einstein's 1907 rejection of the possibility of a special relativistic gravitation theory appears to follow from the inappropriate choice of a scalar gravitational potential. A number of years later Einstein generalised his argument to include all special relativistic gravitation theories which had either scalar or vector potentials. I discuss this generalised argument in the next section.

3.6.2 Einstein's 1912 Force Law Argument

In 1912, Einstein became involved in a vitriolic dispute with Abraham over the question of whether his new gravitation theory amounted to the rejection of the special theory of relativity. [25] Within this dispute, Einstein argued for the impossibility of accounting for gravitation within the framework of special relativity. Einstein's argument contains the essence of his 1907 objection to the possibility of a special relativistic gravitation theory. But it is now in a form that is greatly simplified and also generalised to cover the case of a vector theory of gravitation. He wrote:

> "One of the most important results of relativity theory is the knowledge that all energy E possess inertia (E/c^2) proportional to it. Since every inertial mass is at the same time a gravitational mass, as far as our experience goes, we cannot but attribute to all energy E a gravitational mass E/c^2 also. From this it follows immediately that gravitation acts more strongly on a moving body, than on the same body, in the case that it is at rest.
>
> "If gravitational fields can be explained in the sense of our present day relativity theory, then this can probably only happen in two ways. One can regard the gravitation vector either as a four-vector or as a six-vector. For each of these two cases there are transformation formulae for the transition to a uniformly moving reference system. By means of these transformation formulae and the transformation formulae for pondermotive forces, one can find then the forces acting on moving material points in a static gravitational field for both cases. One comes thereby however to results which contradict the abovementioned consequences of the law of gravitational mass of energy. Therefore it appears that the gravitation vector cannot be fitted into the schema of present day relativity theory without contradiction." [26]

---

25. See note 13, Section 3.4

26. "Eines der wichtigsten Resultate der Relativitätstheorie ist die Erkenntnis, das jegliche Energie E eine ihr proportionale Trägheit (E/c^2) besitzt. Da nun jede träge Masse zugleich eine schwere Masse ist, soweit unsere Erfahrung reicht, können wir

nicht umhin, einer jeden Energie E auch eine schwere Masse E/c^2 zuzuschreiben. Hieraus folgt sofort, dass die Schwere auf einen bewegten Körper stärker wirkt, als auf denselben Körper, falls dieser ruht.

Wenn sich das Schwerefeld im Sinne unserer heutigen Relativitätstheorie deuten lässt, so kann dies wohl nur auf zwei Arten geschehen. Man kann den Gravitationsvektor entweder als Viererverktor oder als Sechservektor auffassen. Für jeden dieser beiden Fälle ergeben sich Transformationsformeln für den Übergang zu einem gleichförmig bewegten Bezugssytem. Mittels dieser Transformationsformeln und der Transformationsformeln für die ponderomotorischen Kräfte gelingt es dann, für beide Fälle die auf in einem statischen Schwerefeld bewegte materielle Punkte wirkenden Kräfte zu finden. Man kommt hierbei aber zu Ergebnissen, die den genannten Konsequenzen aus dem Satz von der schweren Masse der Energie widerstreiten. Es scheint also, dass der Gravitationsvektor sich in das Schema der heutigen Relativitätstheorie nicht widerspruchsfrei einordnen lässt." See
A. Einstein, "Relativität und Gravitation..."pp.1062-3. See also E. Zahar, "Einstein, Meyerson and the Role of Mathematics in Physical Discovery", British Journal for the Philosophy of Science, 31 (1980), pp.20.-21.

Here Einstein refers to two ways of accounting for gravitation within special relativity, as a four-vector and a six-vector. These vectors would appear in the Lorentz covariant force laws

$$F_{\mu} = m\, G_{\mu} \tag{3.19a}$$

and

$$F_{\mu} = m\, U^{\alpha} G_{\alpha\mu} \tag{3.19b}$$

where F_{μ} is the four-force acting on a body of rest mass m and G_{μ} and $G_{\alpha\mu}$ represent the gravitational four- and six- vectors respectively. The four vector equation would typically stem from a scalar theory; G_{μ} would be the $\varphi_{,\mu}$ of equation (3.8a) or some similar term such as in equation (3.8d). The six-vector equation would typically stem from a vector theory; $G_{\alpha\mu}$ would most naturally be the $[\varphi_{\alpha,\mu} - \varphi_{\mu,\alpha}]$ of equation (3.6). In general one would expect $G_{\alpha\mu}$ to be any second rank tensor. The fact that $G_{\alpha\mu}$ must be a six-vector, that is an anti-symmetric second rank tensor, follows from the fact that equation (3.19b) must satisfy the identity $F_{\mu} U^{\mu} \equiv 0$ or in other words $U^{\alpha} U^{\mu} G_{\alpha\mu} \equiv 0$, if m does not vary with φ_{μ}. From this it follows that $G_{\alpha\mu} = -G_{\mu\alpha}$ and the tensor is antisymmetric.

This quote is of great interest because in it we find the most complete statement available of how we can derive the contradiction which Einstein believed faced special relativistic gravitation theories. Einstein's mention of four- and six-vectors suggests that the calculation he outlines is to be carried out in a four dimensional formalism, with force laws similar to (3.19a) and (3.19b). However, if we make this assumption, then it becomes very difficult to determine the exact calculation to which Einstein was referring. Einstein's calculation uses the formulae for transforming quantities from one inertial reference frame to another. But, as I will show below, the contradiction with the "law of the gravitational mass of energy" can be established immediately without recourse to such transformations. Further, if such transformations are to be carried out, it is difficult to envisage how they could be of any special use, for the characteristic property of these four dimensional formulations is that their equations have the same form in all inertial frames of reference.

The simplest explanation of this is that, at the time in question, Einstein was not using the then recently developed four dimensional formalism in his work on gravitation theory. This is consistent with Einstein's early suspicion of mathematics[27] and, in particular, with the fact that his 1912 scalar theories of gravitation were not written within such a four dimensional formalism. Rather he used essentially the same techniques as those used in his 1905 relativity paper. This situation was, however, to change dramatically. In this same year Einstein began work on his generally covariant gravitation theory a theory which demanded the utilisation of an even more sophisticated four dimensional formalism than the one developed by Minkowski.

This suggests that at the time in question, Einstein used the terms four- and six-vector as convenient labels, each denoting a particular gravitation theory, or class of such theories, and that he dealt with them in the older three dimensional formalism. Under these circumstances, the "six-vector" theory would clearly refer to a vector gravitation theory modelled exactly after electrodynamics, with rest mass replacing charge and sign changes where necessary to ensure that like gravitational "charges" attract. The "four-vector"

---

27. R. McCormmach, "Editor's Foreward" in Historical Studies in the Physical Sciences, 7 (1976), (xi)-(xxxv).

theory would be the scalar gravitation theory or theories dealt with earlier.

Under these conditions, we can readily reconstruct the calculations which Einstein described in the quote above and understand why it was necessary to introduce transformation formulae. The calculations are the ones leading up to the scalar gravitation force law (3.17b) and to the Lorentz force law (3.17a), as given earlier. (Of course, in the case of the latter calculation the modification described must be made to convert the theory into a gravitation theory.) In each case, we consider a body stationary in a gravitational field. In the scalar theory case, we require the field to be static. We then transform to a uniformly moving frame of reference to determine the force acting on a moving body. Hence the need for transformation formulae. In the case of the four-vector theory, we find that the quantity corresponding to gravitational mass is $m\sqrt{1-v^2/c^2}$ whilst in the six-vector case it can be seen to be m. The "law of the gravitational mass of energy", however, would require it to be $m/\sqrt{1-v^2/c^2}$. Thus Einstein's contradiction is established.

It is possible to present Einstein's argument in a more general form within a more powerful four dimensional formalism. The results of such an analysis are contained in Table 3.1 below. The calculations supporting these results are given in Appendix A. I consider a static gravitational field, which acts in the x-direction only. Within this field, there is a body of rest mass m moving with instantaneous velocity v in the y-direction alone. F_1, the x-component of the four-force, is calculated for this special case and from it f_1, the x-component of the three dimensional force, using $f_1 = \sqrt{1-v^2/c^2}\,F_1$

From these results, the quantity which would represent "gravitational mass" and the instantaneous acceleration of the body in the x-direction are determined.

For comparison, a corresponding calculation is carried out for a gravitation theory with a 2-tensor potential, $\varphi_{\alpha\beta}$, and thus a 3-tensor gravitation vector, $G_{\alpha\beta\mu}$. The gravitation vectors, G_{μ} , $G_{\alpha\mu}$ and $G_{\alpha\beta\mu}$, are composed solely of combinations of the relevant gravitational potentials and their derivatives, $\varphi_{,\mu}$, $\varphi_{\alpha,\mu}$ $\varphi_{\alpha\beta,\mu}$.

TABLE 3.1 SCALAR, VECTOR AND 2-TENSOR FORCE LAWS

General force law	$F_{\mu} = mG_{\mu}$ (3.19a)	$F_{\mu} = mU^{\alpha}G_{\alpha\mu}$ (3.19b)	$F_{\mu} = mU^{\alpha}U^{\beta}G_{\alpha\beta\mu}$ (3.19c)
Type of gravitation theory	Scalar ("Four-vector")	Vector ("Six-vector")	2-tensor
F_1 for above case	$F_1 = mG_1$	$F_1 = mU^0G_{01}$	$F_1 = mU^0U^0G_{001}$
f_1 for above case	$f_1 = -m\sqrt{1-\frac{v^2}{c^2}}\,G_1$	$f_1 = -mG_{01}$	$f_1 = \frac{-m}{\sqrt{1-\frac{v^2}{c^2}}}\,G_{001}$
"Gravitational mass"	$m\sqrt{1-\frac{v^2}{c^2}}$	m	$\frac{m}{\sqrt{1-\frac{v^2}{c^2}}}$
x-acceleration for above case	$\frac{dx}{dt} = -(1-\frac{v^2}{c^2})\,G_1$ (3.20a)	$\frac{dx}{dt} = -\sqrt{1-\frac{v^2}{c^2}}\,G_{01}$ (3.20b)	$\frac{dx}{dt} = -G_{001}$ (3.20c)

These results match and extend those obtained earlier. We can see that both scalar and vector gravitation theories - theories in which gravitation is represented by a four-vector and six-vector - do not satisfy the requirements of the "law of gravitational mass of energy". In both cases, this failure manifests itself most clearly in the prediction that the downward acceleration of a falling body will be reduced by the body's sideways velocity.

In this sense, Einstein was right when he concluded that gravitation "...cannot be fitted into the schema of present day relativity theory without contradiction". For Einstein's "present day relativity theory" was not the special relativity theory of today. At that time, relativity theory had only dealt thoroughly with a vector field theory, electromagnetism, and to a lesser extent the scalar field theory of gravitation. Einstein had now argued that neither of these two cases could adequately deal with gravitation. Clearly an extension was necessary.

However, Einstein argued that this was to be achieved through a modification of one of the fundamental postulates of special relativity, the light postulate. We can now see that this does not follow directly. If Einstein had considered the case of the second rank tensor theory as analysed in Table 3.1, he would have found that this theory satisfies the requirements of the "law of gravitational mass of energy" and also predicts that the downward acceleration of a falling body is unaffected by its sideways velocity.

Such an anlysis is easy for us to carry out in retrospect and with the advantage of a slightly more flexible formalism. For Einstein, however, this was not the case. Even within the four dimensional formalism of Minkowski, a formalism which Einstein was slowly coming to use, the introduction of a second rank tensor field theory would have been an innovation.

But it cannot be concluded that if Einstein had been made aware of this possibility that he would have abandoned attempts to modify the light postulate as the basis of a gravitation theory. For we have seen that after his work in 1907 led him to suspect that there might be no adequate special relativistic gravitation theory, he came upon the "happiest thought of my life", the equivalence of gravitation and acceleration. The results which followed from this suggested to Einstein that no acceptable gravitation theory was possible within special relativity.

3.6.3 The Argument from the Equivalence of Acceleration and Gravitation

Very soon after he commenced work on the problem of gravitation in relativity theory, Einstein became convinced that the key to the problem lay in the postulated equivalence of acceleration and gravitation. Einstein found that the postulation of this equivalence led to the failure of the light postulate in the presence of a gravitational field and thus the need to modify or extend special relativity theory. This argument is so simple and compelling that it remains even today as one of the most popular arguments for the need of an extension of special relativity to deal with gravitation.

Einstein gives a clear statement of this argument in a 1914 paper in which he gives a pedagogic-historical reconstruction of the origins of relativity theory. After a detailed discussion of the need to extend the principle of relativity to accelerated motion, he describes how he turned to gravitation theory, expecting it to play a fundamental role in this. He continues:

> "On the other hand the fact of experience that all bodies fall equally fast in a gravitational field suggests the understanding that physical processes happen in a gravitational field exactly as they do relative to an accelerated reference system (Equivalence hypothesis). While I laid this understanding as a foundation, I came to the result that

the speed of light should not be considered as independent from the gravitational potential. The principle of the constancy of the speed of light is therefore incompatible with the equivalence hypothesis; relativity theory in the more restricted sense cannot thence be brought into accord with it. Thus I was led to consider relativity theory in the more restricted sense holding only for regions inside of which there are no observable differences of the gravitational potentials. Relativity theory (in the more restricted sense) was to be replaced by a more general theory, which contained within itself the first theory as a boundary case." 28

The simplicity of this argument is deceptive. There are a number of complications that have to be considered. First, it does seem possible for a gravitational disturbance to the propagation of light to be contained within a special relativistic gravitation theory provided we look beyond the range of theories originally considered by Einstein. This can be seen from an examination of the results of Table 3.1. Consider the limiting case that arises

---

28. "Anderseits legt die Erfahrungstatsache, dass alle Körper in einem Gravitationsfelde gleich schnell fallen, die Aufassung nahe, dass sich in einem Gravitationsfelde die physikalischen Vorgänge genau so abspielen, wie relativ zu einem beschleunigten Bezugssytem (Äquivalenzhypothese). Indem ich diese Aufassung zugrunde legte, kam ich zu dem Ergebnis, dass die Geschwindigkeit des Lichhtes nicht als vom Gravitationspotential unabhängig anzusehen sei. Das Prinzip von der Konstanz der Lichtgeschwindigkeit ist also mit der Äquivalenzhypothese unvereinbar; die Relativitätstheorie im engeren Sinne lässt sich daher nicht mit ihr in Einklang bringen. Ich wurde so dazu geführt, die Relativitätstheorie im engeren Sinne nur für Gebiete als zutreffend anzusehen, innerhalb welcher keine merkbaren Differenzen des Gravitationspotentials vorkommen. Die Relativitätstheorie (im engeren Sinne) was durch eine allgemeinere Theorie zu ersetzen, welche erstere als Grenzfall in sich schliesst."

A. Einstein, "Zum Relativitäts-problem", Scientia, 15 (1914), 347.

when the body in question is a photon. That is, we allow the rest mass of the body m to approach zero and the speed of the body v to approach the speed of light c, whilst the energy of the body $\frac{m}{\sqrt{1-v^2/c^2}}$ remains a constant. It now follows from equations (3.20a), (3.20b) and (3.20c) that the photon will not be deflected in the case of the scalar and vector theories, but will be in the case of the tensor theory. This result is plausible physically since in the scalar and vector theories, gravitational mass is based on the photon's rest mass, which is zero. On the other hand, in the tensor theory, gravitational mass is based on the total energy of the photon.

However, it does not follow from this result that Einstein was unjustified in seeking an extension of special relativity theory. The acceptance of this result depends on one's approach to the foundations of special relativity theory. If one takes the attitude of Einstein's 1905 relativity paper and insists dogmatically on a simple operationalist definition of space and time coordinates, as given by the light signalling method, then one cannot accept *ex-hypothesi* that the gravitational bending of light could occur within a special relativistic theory. Such a gravitation theory must automatically involve an extension of special relativity.

On the other hand, one can approach special relativity as stating that spacetime has its own structure, a Lorentzian structure, and that, as an additional empirical result, light is know to propagate along the null geodesics. Under such an interpretation it would seem quite admissible for light propagation to deviate from these null geodesics in the presence of a gravitational field, without any disturbance to the underlying structure of spacetime, or the need for a modification or extension to the special theory of relativity.

However, this in turn leads to a new problem. The action of a gravitational field affects all matter, not just light. This means that in the presence of a gravitational field it is no longer possible to lay out space and time coordinates by the usual simple operational methods. Corrections drawn from theory will always need to appear. For example, one could no longer use a repeatedly reflected light pulse as a clock, unless corrections were made for the effect of the gravitational field on the propagation of the light pulse. Thus the Lorentzian structure of spacetime would become a kind of a ding an sich of which we would have no direct experience.

The acceptance of a special relativistic gravitation theory would therefore depend in part on whether one was willing to accept this state of affairs, in direct contradiction to the approach of Einstein's 1905 relativity paper. It is interesting that Einstein came to accept exactly this state of affairs in the course of the later development of his 1907 speculations on gravitation. In fact, it was to form an essential part of his solution of the problem of gravitation and the general relativity of motion, for it was through the denial of the direct metric significance of space and time co-ordinates that he was able to construct a generally covariant formalism for the writing of physical laws.

But in 1907 Einstein had not yet freed himself from this idea nor had he considered the possibility of a second rank tensor gravitation theory. Even if he had, it had already become unlikely that he would be satisfied with a special relativistic theory of gravitation, for he had become convinced that special relativity was incomplete. The relativity of motion had to be extended from inertial motion to all motion and he believed that the means of carrying this out lay in gravitation theory and in particular in the postulated equivalence of gravitation and acceleration.

3.7 Conclusion

At the beginning of this century, it was becoming apparent that the great advances in the theory of electricity and magnetism of the previous century were to bear unexpected fruit. It was found that these new discoveries could only be satisfactorily integrated into the existing structure of physics if significant changes were made to one of the most important foundations of that structure, the theory of space and time. A new order was asserting itself in the realm of physical theory.

The immediate task at hand was the consolidation of this new order. It was to this end that Einstein's contemporaries approached the problem of gravitation in relativity theory. Their purpose was not to test whether a more profound understanding of gravitation was now accessible to them through the innovations of relativity theory. Rather they sought to show that the hegemony of the new theory of space and time was under no threat from gravitation theory.

But at the same time, in a remarkable display of audacity and flexibility, Einstein was to come to exactly the opposite conclusion. In 1907 he tackled the problem of gravitation in special relativity and rapidly concluded that no acceptable special relativistic gravitation theory was possible. This followed from the failure of such theories to entail the result that all bodies in a given gravitational field fall with the same acceleration, independent of their energy content and sideways velocity. Gravitation theory, which was coming to be regarded as a stagnant area of research, was, in Einstein's view, to strain and burst the framework of the new order, within two years of its erection.

In retrospect, it is easy to see the incompleteness of Einstein's argument. He considered only scalar and vector gravitation theories, which automatically take a source term based on rest mass rather than inertial mass. The choice of inertial mass as a source term leads directly to a second rank tensor gravitation theory. If Einstein had considered such a theory he would have found that it satisfied the above demands: the equality of gravitational and inertial mass, the required independence of acceleration of falling bodies and even the gravitational bending of light.

But Einstein's expectations were swept well beyond this by a far grander vision. Out of electromagnetic field theory had come the first relativistic revolution. Einstein now felt that this revolution was incomplete. He was convinced that the principle of relativity had to be extended from inertial motion to encompass all motion and that it was out of gravitation field theory that this second relativistic revolution was to come. With this end in mind, Einstein elevated the source of the apparent failure of special relativistic gravitation theories to a fundamental principle, the equivalence of gravitation and acceleration and commenced work on a gravitation theory which would satisfactorily embody his new insight.

CHAPTER 4

THE PRINCIPLE OF EQUIVALENCE I

4. THE PRINCIPLE OF EQUIVALENCE I

In 1907 Einstein was struck by a fundamental property of gravitational fields, which had hitherto remained uninterpreted: the motion of a freely falling body is independent of the nature of the body. He guessed that this property was of the greatest significance. On the strength of it he postulated the equivalence of acceleration and gravitation and proceeded to construct a relativistic gravitation theory, which he hoped would ultimately enable the extension of the principle of relativity to include accelerated motion as well.

In this chapter I will map out how Einstein sought to exploit this postulated equivalence of acceleration and gravitation in his gravitation theories prior to 1913. I then review how Einstein sought to carry this equivalence over to his general theory of relativity and the subtle but significant changes which resulted from this.

4.1 The Extension of the Principle of Relativity

In the introduction to his 1907 Jahrbuch... review article of relativity theory, Einstein foreshadowed the concluding sections, which dealt with his speculations on gravitation, with the cautious remark:

> "The further question forces itself whether the principle of relativity is restricted to non-accelerated moving systems. In order not to leave these questions completely out of the discussion, I have added in this essay a fifth part, which includes a new relativistic-theoretic view on acceleration and gravitation." [1]

---

1. A. Einstein, "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen" Jahrbuch der Radioaktivität und Elektronik, 4(1907), 411-62; 5(1908), 9809. Translated by H. M. Schwartz, "Einstein's Comprehensive 1907 Essay on Relativity", American Journal of Physics, 45(1977), 512-7,811-7,899-902. See Schwartz, p.513.

Towards the end of the article, Einstein introduced the sections on gravitation with a similar cautious note:

> "Until now we have applied the principle of relativity - i.e., the assumption that the laws of nature are independent of the state of motion of the reference system - only to nonaccelerated reference systems. Is it conceivable that the principle of relativity holds also for systems which are accelerated with respect to each other?
>
> This is not really the place for the exhaustive treatment of this subject. Since it forces itself, however, on the mind of anyone who has followed the previous applications of the principle of relativity, I shall not refrain here from taking a position on the question..." [2]

With this, Einstein firmly stated the ultimate purpose of his speculations on gravitation - the extension of the principle of relativity to accelerated motion. He also made it quite clear that his treatment of the question was provision, perhaps the beginnings of a lengthier research programme. He then continued to map out the idea which would dominate this programme, the equivalence of acceleration and gravitation:

> "...We consider two system of motion, Σ_1 and Σ_2. Suppose Σ_1 is accelerated in the direction of its X axis, and γ is the magnitude (constant in time) of this acceleration. Suppose Σ_2 is at rest, but situated in a homogenous gravitational field, which imparts to all objects an acceleration $-\gamma$ in the direction of the X axis.
>
> As far as we know, the physical laws with respect to Σ_1 do not differ from those with respect to Σ_2; this derives from the fact that all bodies are accelerated alike in the gravitational field. We have therefore no reason to suppose in the present state of our experience that the systems Σ_1 and Σ_2 differ in any way, and will therefore assume in what follows the complete physical equivalence of the gravitational field and the corresponding acceleration of the reference system..." [3]

---

2. Schwartz, p.899. Einstein's emphasis.

3. Ibid.

In a remarkable autobiographical recollection dating from around 1919, Einstein described how he came to relate the equivalence of acceleration and gravitation to the problem of extending the principle of relativity and the effect it had on him. These recollections are contained in a manuscript at the Einstein Archive, Princeton. The relevant extract has been translated and published by Holton and I have already quoted some parts of it earlier. I now quote Holton's translation at greater length:

> "(15) The fundamental idea of general relativity theory in its original form. In the construction of special relativity theory, the following, [in the earlier part of this manuscript] not-yet-mentioned thought concerning the Faraday [experiment] on electromagnetic induction played for me a leading role.
>
> "According to Faraday, during the relative motion of a magnet with respect to a conducting circuit, an electric current is induced in the latter. It is all the same whether the magnet is moved or the conductor; only the relative motion counts, according to the Maxwell-Lorentz theory. However, the theoretical interpretation of the phenomenon in these two cases is quite different:
>
> "If it is the magnet that moves, there exists in space a magnetic field that changes with time and which, according to Maxwell, generates closed lines of electric force - that is, a physically real electric field; this electric field sets into motion movable electric masses [e.g. electrons] inside the conductor.
>
> "However, if the magnet is at rest and the conducting circuit moves, no electric field is generated; the current in the conductor arises because the electric bodies moving with the conductor experience an electromotive force, as introduced hypothetically by Lorentz, on account of their (mechanically enforced) motion relative to the magnetic field.
>
> "The thought that one is dealing here with two fundamentally different cases was for me unbearable [war mir unerträglich]. The difference between these two cases could not be a real difference, but rather, in my conviction, only a difference in the choice of the reference point. Judged from the magnet there were certainly no electric fields, [whereas] judged from the conducting circuit there certainly was one. The existence of an electric field was therefore a relative one, depending on the state of motion of the coordinate system being used, and a kind of objective reality could be granted only to the electric and magnetic field together, quite apart from the state of relative motion of the observer or the coordinate system. The phenomenon of the electromagnetic induction forced me to postulate the (special) relativity principle.

[Footnote:) the difficulty that had to be overcome [then] was in the constancy of the velocity of light in vacuum which I had first thought I would have to give up. Only after groping for years did I notice that the difficulty rests on the arbitrariness of the kinematical fundamental concepts [presumably such concepts as simultaneity].

"When, in the year 1907, I was working on a summary essay concerning the special theory of relativity for the Jahrbuch für Radioaktivität und Elektronik, I had to try to modify Newton's theory of gravitation in such a way that it would fit into the theory [of relativity]. Attempts in this direction showed the possibility of carrying out this enterprise, but they did not satisfy me because they had to be supported by hypotheses without physical basis. At that point there came to me the happiest thought of my life, in the following form:

"Just as is the case with the electric field produced by electromagnetic induction, the gravitational field has similarly only a relative existence. For if one considers an observer in free fall, e.g. from the roof of a house, there exists for him during his fall no gravitational field - at least in his immediate vicinity. For if the observer releases any objects they will remain relative to him in a state of rest, or in a state of uniform motion, independent of their particular chemical and physical nature. (In this consideration one must naturally neglect air resistance.) The observer therefore is justified to consider his state as one of "rest."

"The extraordinarily curious, empirical law that all bodies in the same gravitational field fall with the same acceleration received through this consideration at once a deep physical meaning. For if there is even a single thing which falls differently in a gravitational field than do the others, the observer would discern by means of it that he is in a gravitational field, and that he is falling into it. But if such a thing does not exist - as experience has confirmed with great precision - the observer lacks any objective ground to consider himself as falling in a gravitational field. Rather, he has the right to consider his state as that of rest, and his surroundings (with respect to gravitation) as fieldfree.

"The fact of experience concerning the independence of acceleration in free fall with respect to the material is therefore a mighty argument that the postulate of relativity is to be extended to coordinate systems that move non-uniformly relative to one another..." [4]

---

4. A. Einstein, "Fundamental Ideas and Methods of Relativity Theory, Presented in their Development" trans. G. Holton in "Finding Favour with the Angel of the Lord. Notes towards the Psychobiographical Study of Scientific Genius" in Y. Elkana, ed., The Inter action Between Science and Philospghy (Humanities Press, 1975), pp.369-71. Emphasis as in this translation.

Here Einstein recalls his 1905 work on the special theory of relativity. It had stemmed from the discovery of a defect in the then current theory of electrodynamics, which he illustrated with the celebrated example of a magnet and conductor in relative motion. On the theoretical level, electrodynamics distinguished states of absolute rest and motion. A magnet in absolute motion generated an induced electric field in the space around it. A magnet in a state of absolute rest did not. The absolute existence of this induced electric field, which could be observed from any frame of reference, guaranteed that the magnet was in absolute motion. Yet this distinction did not appear in the phenomena themselves, in that which was actually measured. All that seemed to matter was the relative velocities concerned. The measureable current induced in a conductor by its motion with respect to a magnet depended only on their relative velocities and not their absolute velocities.

Einstein eliminated this defect in a new account of electrodynamics in which the existence of electric and magnetic fields was purely relative. The existence of an induced electric field around a magnet was no longer an absolute which depended only on the absolute state of motion of the magnet. Rather the judgement of the existence of such a field depended only on the relative motion of the magnet and the observer. An induced electric field was only observed if there was relative motion between the magnet and observer. The question of the absolute motion of the magnet no longer arose and, with this, the need for a distinction between absolute rest and motion disappeared from electrodynamics.

Continuing in the above extract, Einstein refers to his 1907 work on gravitation and his failure to find an acceptable relativistic gravitation theory. This failure led Einstein to the "happiest thought of my life". He recognised that gravitational fields, just like electric and magnetic fields, have a relative existence. A freely falling observer in a gravitational field will see no field around him. Einstein discussed the complementary case in the extract from the 1907 Jahrbuch... article quoted above. The inertial field produced by the acceleration of an observer is also not an absolute. For the phenomena themselves give the observer no means of distinguishing this inertial field from a gravitational field.

The belief in the absolute status of electric and magnetic fields had lured physicists to introduce the spurious concept of absolute rest. Once again, a belief in the absolute nature of inertial fields had led physicists, in this case since the time of Newton, to a concept of absolute acceleration. Indeed this belief underpinned Newton's well-known bucket thought experiment. [5] The relative nature of inertial and gravitational fields - expressed by his postulation of the equivalence of acceleration and gravitation - was now to be the key to the elimination of the absolute status of acceleration and the extension of the principle of relativity.

Returning to the 1907 Jahrbuch... article, we see that there Einstein proceeded to note that the equivalence of acceleration and gravitation, as postulated for uniform acceleration and homogeneous

---

5. I. Newton, Principia, trans. A. Motte., rev. F. Cajori (Berkeley: University of California Press, 1962). pp.10-11.

gravitational fields, enabled a limited extension of the principle of relativity. He then noted how the equivalence was going to be used by him, presumably in his attempts to effect a more complete extension:

> "...This assumption extends the principle of relativity to the case of uniformly accelerated translational motion of the coordinate system. The heuristic value of the assumption lies therein that it makes possible the replacement of a homogeneous gravitational field by a uniformly accelerated reference system, the latter case being amenable to theoretical treatment to a certain degree." [6]

Then Einstein went on to describe a new theory of acceleration and gravitation, which exploited this equivalence to the full.

4.2 The 1907 Jahrbuch Theory

The theory which Einstein proceeded to describe began with a reanalysis of the kinematics of acceleration. He considered a uniformly accelerating frame of reference and determined the behaviour of rods and clocks in this frame, for the special case of small accelerations and regions of small spatial extension, by a simple application of the results of his special theory of relativity. He concluded that the length of rods in all three coordinate directions will be unaffected by the acceleration, since such an effect must be an even function of acceleration and the acceleration has been assumed to be small. However, the speed at which clocks run was found to vary with direction in the accelerating frame. For acceleration γ in spatial coordinate direction ξ the time read by a clock, σ, is given by $\sigma = \tau\left[1 + (\gamma\xi/c^2)\right]$ (4.1a)

6. Schwartz, p.899.

where τ is the time read by a clock at the origin of coordinates.

Einstein then turned to gravitation. From the equivalence of uniform acceleration and homogeneous gravitational fields, it followed that this result must hold for a homogeneous gravitation field, with gravitational potential $\Phi = \gamma \xi$,

$$\sigma = \tau[1 + \Phi/c^2] \tag{4.1b}$$

and the other symbols have the same meaning as above. Extending this result to the nonhomogeneous gravitation field of the sun, he was able to predict a small red shift in light coming from the sun.

If we now recall the structure of Einstein's 1905 paper on special relativity, we find an interesting parallel.[7] Einstein began by recognising a threat to the equivalence of inertial frames of reference in the then current theory of electrodynamics. He met the threat in a revolutionary way. He decided that the problem did not lie in electrodynamics, but in kinematics. On the basis of the reassertion of the equivalence of inertial frames of reference, in conjunction with the light postulate, he developed a new kinematics of inertial motion, in the first "Kinematical Part" of the paper. With this new kinematics, he returned to electrodynamics, in the second "Electrodynamical Part", and showed how its theoretical structure no longer required a preferred state of motion to be set aside. In particular, he demonstrated how the basic equations of electrodynamics could remain form invariant under transformations between inertial frames of reference.

---

7. A. Einstein,"On the Electrodynamics of Moving Bodies" in The Principle of Relativity (New York: Dover, 1952), pp.35-65.

In 1907 Einstein found himself faced with a similar problem. We have already seen how he drew a parallel between the two cases. In 1905, the presumed absolute nature of electric and magnetic fields had threatened the equivalence of inertial frames of reference. Once again the presumed absolute nature of inertial forces compromised the equivalence of accelerated and unaccelerated frames of reference. In his 1907 reply, Einstein seemed to be using the same tactic. Once again he was shifting the locus of the problem away from the mechanics of inertial forces to the kinematics of acceleration and gravitation. Once again he was developing a new kinematics, but this time of acceleration.

Now, if Einstein were actually following the general model of his 1905 special relativity paper, then we would expect him to continue by applying this new kinematics to other areas, to the question of gravitational and inertial forces in particular, and show how it enabled the equivalence of accelerated and unaccelerated frames of reference to be realised.

This he did, but only to a very limited extent. First he applied the new kinematics to electromagnetism. He transformed Maxwell's equations to a uniformly accelerated frame of reference and invoked the equivalence of acceleration and gravitation. Just as was the case in his 1905 paper, he found that he could retain the form of the equations in all the reference frames considered, if he adopted certain transformation laws for the quantities they dealt with. One of these involved the speed of light. This speed was to be given by the expression

$$c\left[1+(\gamma\xi/c^2)\right] = c\left[1+\Phi/c^2\right]$$

where c is the speed of light in a gravitation and acceleration free frame of reference. This in turn entailed the bending of light rays by a gravitational field. But, because of the weakness of the earth's gravitational field, Einstein believed that there was no prospect of an experimental test of this prediction.

Out of this, Einstein produced one last result of great interest. He contracted and integrated Maxwell's equations in order to yield a law of conservation of energy. From this he found that to every quantity of energy E, he must attribute an extra energy $\frac{E\Phi}{c^2}$. From this he concluded that to every quantity of energy E there corresponds not only an inertial mass E/c^2, but also a gravitational mass E/c^2. With this result, we see the beginnings of the applicaton of the new kinematics to the mechanics of inertial forces. For out of the application of the new kinematics to Maxwell's equations, Einstein has been able to recover a crucial result in the mechanics of inertia and gravitation, the equality of inertial and gravitational mass.

The theory developed here was far from complete, as Einstein had made clear in his introduction. But within it we can discern the outline of a programme through which the extension of the principle of relativity could be realised. It would be a two stage affair. The first step involved the construction of a new kinematics of acceleration and gravitation, just as his 1905 work had first required the construction of a new kinematics of inertial motion. Then this new kinematics would be mated with the remainder of the theories of physics, most probably by extending the form invariance of their basic equations to the case of accelerating frames of reference and those under the influence of a gravitational field.

4.3 The 1911 Annalen der Physik Arguments

Einstein did not publish any further in this area until his well known 1911 paper, "On the Influence of Gravitation on the Propagation of Light", in the Annalen der Physik.[8] Superficially, the content of the paper was little different to his 1907 work. In the introduction to the paper, he said that he was returning to the question of the influence of gravitation on the propagation of light because he was dissatisfied with his earlier presentation of the subject and "for a stronger reason": he now saw that the predicted bending of a ray of light by a gravitational field could be tested observationally by looking for shifts in the apparent position of stars whose light grazes the sun on its journey to earth.

The only new result of the paper was the calculation of the angle of deflection of the apparent position of such a star, 0.83 seconds of arc. The basic method of the paper was the same as in 1907: the effect of uniform acceleration on a process was determined and then assumed to be identical with that of a homogeneous gravitational field. The effects recovered were essentially the same: the gravitational retardation of clocks, the gravitational red shifting of light from the sun, the bending of light and the equality of gravitational and inertial mass.

---

8. In The Principle of Relativity (New York: Dover, 1952), pp.99-108.

But there were important differences between this paper and Einstein's 1907 presentation. First, Einstein's hope that this developing gravitation theory would lead to an extension of the principle of relativity was not clearly stated. In half a sentence, he mentioned that the exact equivalence of gravitation and acceleration made it "impossible for us to speak of the absolute acceleration of the system of reference, just as the usual theory of relativity forbids us to talk of the absolute velocity of a system".[9] But this was all he said and it was buried in the body of the paper. The purpose of the paper seemed to be solely to develop the implications of his gravitation theory for processes in a gravitational field.

Further, the structure of the argument had changed. The 1907 presentation had followed the layout of Einstein's 1905 relativity paper. First he had dealt with a purely kinematical problem, the behaviour of accelerated clocks and rulers, and then used the results of this analysis with his gravitation theory to determine the effects of gravitation on various processes. In the 1911 paper, Einstein dispensed with this division and, in the process, he was able to simplify and generalise his arguments.

In the 1907 presentation, the equality of gravitational and inertial mass had been derived from Maxwell's equations in the presence of a gravitational field and for electromagnetic energy only. In the 1911 paper, Einstein tackled this question first. He produced a two step argument. First he considered the Doppler shifting of the energy of light propagating in an accelerating frame

---

9. Ibid., p.100.

of reference and, invoking the equivalence of gravitation and acceleration, argued that we must attribute the extra gravitational energy $\frac{E\Phi}{c^2}$ to any quantity of light energy E in a gravitational field. He then considered a cyclic series of operations involving the transmission of light energy and the raising and lowering of a mass in a gravitational field. The application of the requirement of the conservation of energy to the cycle led directly to the result that any inertial mass E/c^2 has an equal gravitational mass E/c^2. This result was clearly more general than the 1907 result.

Einstein continued to consider a more compact derivation of the same result. He considered bodies suspended on springs in gravitational and inertial fields. The extension of the spring indicated the gravitational and inertial mass in each case respectively. From the equivalence of gravitation and acceleration and the inertia of energy, it followed that an increase in the energy of a body E also increased its inertial mass by E/c^2 and its gravitational mass by an equal amount E/c^2.

Einstein completed the paper by deriving the gravitational red shift of light from the Doppler shift of the frequency of a light signal in an accelerated frame of reference. This result led him in turn to the gravitational retardation of clocks, the variation of the speed of light with gravitational potential and finally to the deflection of a ray of starlight by the sun.

The 1911 paper is a paper which is remarkable for the power and compactness of the arguments which it contains. But it is a paper which, if seen in isolation, does not give a clear picture of what

Einstein originally hoped to achieve with his gravitation theory. This could be seen more clearly in the 1907 presentation and also in the 1912 papers which Einstein was to publish on his gravitation theory. There Einstein gives a purely kinematical treatment of acceleration and uses the results to show how the form of Maxwell's equations could be retained in an accelerated frame of reference or within a gravitational field. This reflects his original concern, the theory of space and time and the extension of the principle of relativity to accelerated motion.

4.4 The 1912 Scalar Theory of Gravitation

In 1912, Einstein published a series of papers in which he presented work forming the most complete development of his work on gravitation. The first of these, "The Speed of Light and Static Gravitational Fields," contained the beginning of an elaborate extension of the 1907 and 1911 theory of gravitation.[10]

After introductory preliminaries, the paper returned to the purely kinematical analysis of the effects of acceleration of the 1907 paper. The method Einstein used closely parallels the method he used in the special relativistic part of the 1907 *Jahrbuch...*article to determine the effects of uniform velocity and derive the Lorentz transformation. This latter method depended on considering an expanding shell of light from two inertial frames of reference. In conjunction with a number of suitable assumptions, he concluded that the only coordinate transformation which would leave

---

10. A. Einstein, "Lichtgeschwindigkeit und Statik des Gravitationsfeldes", *Annalen der Physik*, *38* (1912), pp.355-69.

the expression describing this process,

$$x^2 + y^2 + z^2 = c^2t^2,$$

form invariant was the Lorentz transformation.

In the 1912 paper, Einstein considered an inertial frame of reference $\Sigma(\xi,\eta,\zeta,\tau)$ in which the speed of light is unity and an accelerated frame K(x,y,z,t) in which the variable speed of light is c. He then sought the transformation equations relating these two frames through consideration of an expanding shell of light which would be described by

$$d\xi^2 + d\eta^2 + d\zeta^2 - d\tau^2 = 0$$

and $$dx^2 + dy^2 + dz^2 - c^2dt^2 = 0$$

in each frame of reference respectively. From this he drew the identity

$$d\xi^2 - d\tau^2 = dx^2 - c^2dt^2 \qquad (4.2)$$

dropping η and ζ
since K accelerates in the ξ direction of Σ only.

On the basis of a number of suitable assumptions, including the assumption that the acceleration of K is small and that t remains small, Einstein arrived at the transformation equations

$$\left.\begin{array}{l} \xi = x + \frac{ac}{2}t^2 \\ \eta = y \quad \zeta = z \quad \tau = ct \end{array}\right\} \qquad (4.3a)$$

where $$c = c_0 + ax \qquad (4.3b)$$

and c_0 and a are constants.

In the remainder of the paper and in a paper which followed immediately after, Einstein turned to the task of using these results to develop the new gravitation theory.[11] Out of equation (4.3b), he developed a field equation for static gravitational fields in which the variable speed of light c figured as the scalar gravitational potential. In the second paper, he found it necessary to return to modify this field equation, for he discovered that it violated the equality of action and reaction. From equations (4.3a) he developed equations of motion for masses in a static gravitational field and he also examined the effect of such a field on the rate of clocks. In later sections and chapters, I will return to examine features of this theory.

In the second paper, Einstein examined the effect of gravitation on other processes. Using familiar equivalence arguments, he established that the form of Maxwell's equations remained essentially the same as in the absence of a gravitational field. The only change was in the insertion of the gravitational potential/ speed of light c in the two "curl" terms. A similar analysis of thermal quantities showed that a minor modification was all that would be required to the equations of thermodynamics in order to account for the effects of a static gravitational field.

These attempts to retain the special relativistic form of the laws governing processes which occur in accelerated frames of reference or in gravitational fields clearly relate to attempts to extend the range of equivalent frames of reference beyond the set of inertial frames of reference. But Einstein carefully avoided detailed discussion of the question of a wider equvalence of frames of

---

11. A. Einstein, "Zur Theorie des statischen Gravitationsfeldes", Annalen der Physik, 38 (1912), pp.443-58.

reference and how this could be achieved by extending the range of frames of reference in which the laws of nature retain their special relativistic forms. He limited his comments to a short passage at the end of the first paper:

> "To me the space-time problem appears to lie in the following. If one restricts oneself to a region of constant gravitational potential, then the laws of Nature take on an especially simple and invariant form, provided one relates them to space-time systems of that manifold, which is connected through the Lorentz transformations with constant c. If one does not limit oneself to regions of constant c, then the manifold of equivalent systems, as well as the manifold of natural laws left unchanged by transformations, becomes greater. But for this, the laws become more complicated." 12

The possibility of extending the "manifold of equivalent systems" is, of course, exactly what is required by an extension of the principle of relativity. But the possibility of this happening seems little closer than it did in 1907. Whilst Einstein had added considerable detail to his gravitation theory, there had been no qualitative advance as far as extending the principle of relativity was concerned since 1907. Einstein no longer even mentioned an extension of the principle explicitly.

---

12. "Mir scheint das Raum-Zeitproblem wie folgt zu liegen. Beschränkt man sich auf ein Gebiet von konstantem Gravitationspotential, so werden die Naturgesetze von ausgezeichnet einfacher und invarianter Form, wenn man sie auf ein Raum-Zeitsystem derjenigen Mannigfaltigkeit bezieht, welche durch die Lorentztransormationen mit constantem c miteinander verknüpft sind. Beschränkt man sich nicht auf Gebiete von konstantem c so wird die Mannigfaltigkeit der äquivalenten Systeme, sowie die Mannigfaltigkeit der die Naturgesetze ungeändert lassenden Transformationen eine grössere werden, aber es werden dafür die Gesetze komplizierter werden."

    Einstein, "Lichtgeschwindigkeit...", pp.368-9.

But Einstein had not completely lost sight of this goal. In another 1912 paper, in which he defended his gravitation theory from an attack by Abraham, he wrote:

> "On the other hand this principle of equivalence opens for us the interesting perspective, that the equations of a relativity theory, which also includes gravitation, might also be invariant with respect to acceleration (and rotation) transformation. Admittedly the way to this goal seems certainly to be more difficult. One can see already from the highly specialised case of the gravitation of resting masses treated hitherto, that space-time coordinates will forfeit their simple physical meaning, and it is still not forseen what form the general spacetime transformation equations could have. I would like to ask all colleagues to work on this important problem." 13

The search for "general spacetime transformation equations" must be the search for the relations between clock readings and ruler measurements in arbitrary frames of reference. They could not be the general coordinate transformations of the general theory of relativity, for there can be no problem in discovering their form. General covariance demands that they have no special form! This places Einstein's comment well within the 1907 programme. But we also see here a number of features that will figure prominently in the gestating general theory of relativity: the invariance of the form of equations and, in particular, the non-metrical significance of the coordinates.

---

13. "Anderseits eröffnet uns dies Äquivalenzprinzip die interssante Perspektive, dass die Gleichungen einer auch die Gravitation umfassenden Relativitätstheorie auch bezüglich Beschleunigungs- (und Drehungs-) Transformationen invariant sein dürften. Allerdings scheint der Weg zu diesem Ziele ein recht schwieriger zu sein. Man sieht schon aus dem bisher behandelten, höchst speziellen Falle der Gravitation ruhender Massen, dass die Raum-Zeit Koordinaten ihre einfache physikalische Deutung einbüssen werden, und es ist noch nicht abzusehen, welche Form die allgemeinen raumzeitlichen Transformationsgleichungen ghaben könnten. Ich möchte alle Fachgenossen bitten, sich an diesem wichtigen Problem zu versuchen."
A. Einstein, "Relativität und Gravitation", Annalen der Physik, 38(1912), pp.1063-4.

There had been no real advances in his programme for five years and now Einstein was calling for help from his colleagues. What Einstein sorely needed was a new result to confirm his faith in the programme and display its productivity to his colleagues. This discovery came in 1912 and led Einstein to publish again even before the final papers on his scalar gravitation theory had appeared.

4.5 The Origin of Inertia

The paper in question, "Is there a Gravitational Effect which is analogous to electrodynamic Induction?", considered a mass surrounded by a homogeneously distributed spherical shell of matter.[14] Einstein examined the interaction of the central mass and the mass in the shell from the point of view of his 1912 gravitation theory. The urgency with which he published can be judged by the fact that he had to acknowledge that the details of his gravitation theory were still to appear "shortly" in the Annalen der Physik.

The paper contains two important conclusions which relate to the origin of inertia. First he concluded from a fairly simple weak field approximation that the inertial masses of the two bodies in question are each increased by the presence of the other bodies. He described the significance of this result:

> "This result is in itself of great interest. It shows that the presence of the inertial casing K increases the inertial mass of the material point P found within it. This suggests that all the inertia of a point mass is an effect of the existence of all other masses, resting on a kind of interaction with the latter. The extent to which this understanding is justified will show itself if we will be fortunate to have a serviceable

---

14. A. Einstein, "Gibt es eine Gravitationswirkung die der elektromagnetischen Induktionswirkung analog ist?", Vierteljahrsschrift für gerichtliche Medizin, 44 (1912), pp.37-40.

dynamics of gravitation."

In a footnote to the second last sentence, Einstein noted:

> "This is exactly the point of view which E. Mach maintained in his acute investigations on the subject. (E. Mach, The Development of the Principles of Dynamics. Second Chapter. Newton's Views of Time, Space and Motion.)" [15]

Here we find Einstein discussing for the first time an idea which would feature prominently in the development of gravitation theory and the general theory of relativity: that the inertia of a body is due to an interaction with all other bodies. In a later chapter, I will consider the status of this idea and the extent to which it can be attributed to Mach, as Einstein does here.

In order to clarify the argument, Einstein completed the paper by studying how this inertial interaction between bodies comes about. He showed that if the spherical shell accelerates, it will tend to drag the central body along with it. This is, of course, directly analogous to the inductive forces which arise between accelerated charges, except for a change of sign.

---

15. "Das Resultat ist an sich von grossem Interesse. Es zeigt, dass die Anwesenheit der trägen Hülle K die träge Masse des darin befindlichen materiellen Punktes P erhöht. Es legt dies die Vermutung nahe, dass die ganze Trägheit eines Massenpunktes eine Wirkung des Vorhandenseins aller übrigen Massen sei, auf einer Art Wechselwirkung mit den letzteren beruhend. Inwieweit diese Auffassung berechtigt ist, wird sich zeigen, wenn wir in dem glücklichen Besitze eine brauchbaren Dynamik der Gravitation sein werden."

    "Es ist dies ganz derjenige Standpunkt, welchen E. Mach in seinen scharfsinnigen Untersuchungen über den Gegenstand geltend gemacht hat. (E. Mach, Die Entwicklung der Prinzipien der Dynamik. Zweites Kapitel. Newtons Ansichten über Zeit, Raum und Bewegung.)"

    Ibid., p.39.

This at last suggests a more complete account of the origin of inertial forces, which is still in accord with an extended principle of relativity. It suggests that the inertial forces which act on an accelerating body do not arise from an interaction with space, an explanation which presupposes the existence of preferred inertial frames of reference in space. Rather these forces arise because the body is accelerating with respect to a frame of reference in which the bulk of the masses in the universe move at constant velocity. They arise as a part of the gravitational interactions of all masses of the universe, in much the same way as inductive forces arise from the electromagnetic interactions of accelerating charges.

Thus an extension of the principle of relativity at last seems possible and, in accord with Einstein's 1907 expectations, stems from a theory of gravitation and inertia which is itself based on the equivalence of gravitation and acceleration.

It is not clear to what extent Einstein expected such a result on the nature of inertia. to follow from his gravitation theory when he began work on it in 1907. In later recollections he described his 1907 approach to the problem of graviation:

> "I was of course acquainted with Mach's view, according to which it appeared conceivable that what inertial resistance counteracts is not acceleration as such but acceleration with respect to the masses of the other bodies existing in the world. There was something fascinating about this idea to me, but it provided no workable basis for a new theory.
>
> "I first came a step nearer to the solution of the problem when I attempted to deal with the law of gravity within the framework of the special theory of relativity..." [16]

---

16. A. Einstein, "Notes on the Origin of the Genral Theory of Relativity" in Ideas and Opinions (London: Souvenir Press, 1973) p.286.

But there is no mention of "Mach's view" until 1912 in his publications.

It is also interesting to note that had Einstein sought to recover only the inductive effects of the 1912 paper from a gravitation theory, he would have had little trouble. For it seems that just about any field theory, in which there is a maximum speed at which interactions can propagate through the field, can yield such an effect. This is the case, for example, with Maxwell's electrodynamics. So if Einstein had been searching for such an effect since 1907, then something must have delayed him. Perhaps it was the fact that even his 1912 theory dealt only with static fields, or that he first wanted to find a satisfactory gravitation theory based on the equivalence of gravitation and acceleration.

4.6 The Weakening of the Principle of Equivalence

In the course of the year following this last paper, the major breakthrough came. In 1913, in conjunction with the mathematician Marcel Grossmann, Einstein published the "Entwurf..." paper, in which the fundamental techniques of tensor calculus were described and the basic structure of the general theory of relativity laid down.[17] This step marked a turning point for the principle of equivalence*, for within the general theory of relativity the content

---

* In the course of 1912 and 1913, Einstein began to introduce the terms _Äquivalenzprinzip_ and _Äquivalenzhypothese_ as a shorthand label for his hypothesis of the equivalence of acceleration and gravitation. The former eventually came to be used exclusively. Henceforth, I shall label the hypothesis and its later variants the "principle of equivalence", unless there is danger of confusion.

17. A. Einstein & M. Grossmann, "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation", _Zeitschrift für Mathematik und Physik_, 62 (1913), pp.225-261.

and import of the principle came to be changed and weakened. In this section, I describe the nature of this change and seek to map out the function of the principle in the new theory.

4.6.1 The Scalar Theory

An examination of the development of Einstein's scalar gravitation theory in the 1907-1912 period shows that this weakening process had already begun then. Einstein's statement of the equivalence remained essentially the same as in the original 1907 paper, as stated in the passage quoted in Section 4.1 above. Perhaps the only later addition was a careful definition of exactly what was meant by the uniform acceleration of a system of motion. [18]

This equivalence is asserted to hold only between uniform acceleration and homogeneous gravitational fields. Now the heart of Einstein's 1907 hopes for an extension of the principle of relativity to accelerated motion was the idea that gravitation could be seen to be interchangeable with acceleration. This suggests that a wider equivalence of acceleration and gravitation ought to be sought. In fact, however, in 1911 and 1912, Einstein faced difficulties in even retaining this limited equivalence. In a footnote to his 1911 Annalen der Physik paper, he conceded that the equivalence could not obtain globally for arbitrary gravitational fields. [19]

This was certainly not a great concession for Einstein to make. But in 1912, he was forced to make a far more significant one. Whilst he recognised that the equivalence did not hold globally for arbitrary fields, he believed that it did hold globally for at least some cases

---

18. Einstein, "Über das Relativitätsprinzip...", p.99.

19. Einstein, "One the Influence...", p.100.

- specifically homogenous gravitational fields and uniform acceleration. In 1912, he found that the field equation, which seemed to follow directly from this assumption, violated the requirement of the equality of action and reaction. It predicted that gravitational masses, linked together by a rigid massless frame, would set themselves in motion as a whole.[20]

In a later chapter I will examine how Einstein grappled with this problem until he was forced to concede a modification to his field equation. He approached this outcome reluctantly:

> "Therefore I resolved myself to this step with difficulty, for with it I abandon the grounding of the unconditional principle of equivalence. It appears that this can only be maintained now for infinitely small fields."[21]

The degree of Einstein's reluctance can be gauged by the fact that in an otherwise very compact presentation, Einstein devoted considerable space to the task of convincing the reader of the necessity of this step. Elsewhere, Einstein discussed this restriction on the principle of equivalence. He reaffirmed his faith in the principle but conceded that he knew of no satisfactory reason for the restriction.[22]

In spite of this concession, the principle of equivalence still played a central role in Einstein's work on gravitation in the 1907-1912 period. Apart from the field equation, little had to be modified

---

20. Einstein, "Zur Theorie...", p.452.

21. "Zu diesem Schritt entschliesse ich mich deshalb schwer, weil ich mit ihm den Boden des unbedingten Äquivalenzprinzips verlasse. Es scheint, dass sich letzteres nur für unendlich kleine Felder aufrecht erhalten lässt."
<u>Ibid</u>., pp.455-6.

22. "...und dass ich hierfür keinen befriedigenden Grund anzugeben weiss."
Einstein, "Relativität und Gravitation", p.1063.

in the 1912 scalar theory to accommodate the shift in the principle of equivalence from a global principle to a local principle. The function of the principle remained the same. It acted as a bridge between acceleration and gravitation. It enabled the effects of gravitational fields to be determined from an examination of the far more accessible effects of an acceleration of the frame of reference. Time and time again, Einstein applied this method. He wanted to know the effect of gravitation on the propagation of light, on the motion of bodies, on electromagnetic systems. So he would first determine the effects of acceleration on each of these processes and then invoke the principle of equivalence to recover his final result.

4.6.2 The General Theory of Relativity

However, with the emergence of the general theory of relativity, both the content and import of the principle of equivalence was to change significantly. This change was disguised to some extent by Einstein's presentation of the content and significance of the principle in introductions to and summaries of the theory both during and after its perfection in the 1913-1915 period. There he remained faithful to his 1907 vision of the principle of equivalence and its significance to a general theory of relativity. Indeed, even the actual statement of the equivalence remained very similar to the 1907 statement. For example, it appears in the following way in his well known 1916 exposition of the theory:

> "In addition to this weighty argument from the theory of knowledge, there is a well known physical fact which favours an extension of the theory of relativity. Let K be a Galilean system of reference, i.e. a system relatively to which (at least in the four-dimensional region under consideration) a mass, sufficiently distant

from other masses, is moving with uniform motion in a straight line. Let K' be a second system of reference which is moving relatively to K in uniformly accelerated translation. Then, relatively to K', a mass sufficiently distant from other masses would have an accelerated motion such that its acceleration and direction of acceleration are independent of the material composition and physical state of the mass.

"Does this permit an observer at rest relatively to K' to infer that he is on a 'really' accelerated system of reference? The answer is in the negative; for the above-mentioned relation of freely movable masses to K' may be interpreted equally well in the following way. The system of reference K' is unaccelerated, but the space-time territory in question is under the sway of a gravitational field, which generates the accelerated motion of the bodies relatively to K'.

"...Therefore, from the physical standpoint, the assumption readily suggests itself that the systems K and K' may both with equal right be looked upon as 'stationary', that is to say, they have an equal title as systems of reference for the physical description of phenomena." [23]

This statement of the equivalence is essentially identical to Einstein's 1907 version (See Section 4.1) The key idea of both is the indistinguishability of uniform acceleration and homogeneous gravitational fields. For the rest of his life, Einstein usually stated the principle in this form and presented it as a fundamental concept of the general theory of relativity. [24] I will now try to show that this statement of the principle no longer reflected the role of the principle in the theory or its content.

---

23. A. Einstein, "The Foundation of the General Theory of Relativity" (1916), in The Principle of Relativity (New York: Dover, 1952), pp.113-4. Einstein's emphasis.

24. See, for example, one of his last statements of the principle in A. Einstein, "On the Generalized Theory of Gravitation" (1950), in Ideas and Opinions (London: Condor, 1973), pp.346-7.

We have seen that the repeated invocation of the principle of equivalence formed the central part of Einstein's 1907 - 1912 gravitation theories. With the emergence of the "Entwurf..." paper in 1913, this was to change. The principle of equivalence was no longer invoked throughout the development and use of the general theory of relativity; its use was limited to the establishment of one single result, that gravitational fields could be accounted for by a spacetime structure which had a quadratic metric.

Corresponding to this, there was a shift in the content of the principle, although Einstein seemed to avoid acknowledging it. In a 1918 summary, Einstein presented the foundations of the theory as three powerful physical principles: the principle of relativity, the principle of equivalence and Mach's principle. There he gave his most compact and accurate statement of the content of the principle of equivalence and its role in his general theory of relativity:

> "Principle of Equivalence: Inertia and weight are the same in essence. From this and from the results of the special theory of relativity it necessarily follows that the symmetrical 'fundamental tensor' ($g_{\mu\nu}$) determines the metrical properties of space, the inertial behaviour of bodies in it, as well as gravitational action..." [25]

---

25. Äquivalenzprinzip: Trägheit und Schwere sind wesensgleich. Hieraus und aus den Ergebnissen der speziellen Relativitätstheorie folgt notwendig, dass der symmetrische 'Fundamentaltensor' ($g_{\mu\nu}$) die metrischen Eigenschaften des Raumes, das Trägheitsverhalten der Körper in ihm, sowie die Gravitationswirkungen bestimmt..."

A. Einstein, "Prinzipielles zur allgemeinen Relativitätstheorie", Annalen der Physik, 55 (1918), p.241.

The explicit statement of the principle as the understanding that inertia and weight are the same in essence is an innovation of this 1918 paper and, as we shall see, only emerged after some confusion over the content of the principle in the years preceding. However, the role of the principle, as sketched out above, remained the same in Einstein's expositions of his theory from the very earliest in 1913. In all Einstein's major expositions of the current state of his theory in the crucial 1913 - 1916 period, the principle of equivalence contributed to the establishment of the above result in an argument that remained essentially the same. [26]

In general the argument ran as follows. We consider a (possibly infinitely small) region of space from a coordinate system in which special relativity holds. This means that the metric for its space-time structure is given by

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 \qquad (4.4a)$$

and the motion of a free mass point is governed by the action principle

$$\delta\int ds = 0 \qquad (4.5)$$

A general coordinate transformation is now introduced:

$$\begin{aligned} x' &= x'(x,y,z,t) \\ y' &= y'(x,y,z,t) \\ z' &= z'(x,y,z,t) \\ t' &= t'(x,y,z,t) \end{aligned}$$

so that the coordinate differentials transform according to

$$dx^{\mu} = a^{\mu}_{\nu}\, dx'^{\nu}$$

where $(x^0,x^1,x^2,x^3) = (t,x,y,z)$ and a^{μ}_{ν} are determined by form of the transformation.

---

26. See for example A. Einstein, "Zum gegenwärtigen Stande des Gravitationsproblems", Physikalische Zeitschrift, 14 (1913), pp.1255-6; A. Einstein, "Die formale Grundlage der allgemeinen Relativitätstheorie", Preussische Akademie der Wissenschaften, Sitzungsberichte, 1914 Part 2, pp.1032-3; Einstein, "The Foundation of...", pp.118-20.

Substitution into the metric (4.4a) yields

$$ds^2 = g_{\mu\nu} dx'^{\mu} dx'^{\nu} \tag{4.4b}$$

where the $g_{\mu\nu}$ are symmetric in μ and ν and functions of x'^{μ}.

The motion of free mass points in the primed coordinate system is still governed by the action principle (4.5), for the interval ds is a quantity which, by its nature, is independent of coordinate systems.

In general, the transformation to the primed coordinate system will be to an "accelerated" system and therefore free mass points will move non-uniformly in curvilinear paths. At this point in the argument, Einstein refers back to a number of considerations developed before the presentation of the argument. These are the requirements of the general relativity of motion; the equal admissability of all coordinate systems for the writing of the laws of nature; the conclusion that space and time coordinates will lose their direct metrical significance and space its Euclidean nature in such general coordinate systems; and the principle of equivalence. On the basis of the reference to these considerations, Einstein completes the argument with the conclusion that the metric tensor now also characterises gravitational fields, as well as the inertial properties and general structure of spacetime and that the action principle (4.5) governs the motion of a free mass point in a gravitational field.

It is curious that Einstein should need to resort to the wider considerations used to complete the final step of the argument. We have seen that in 1907 Einstein expected that the equivalence of gravitation and acceleration would lead to an extension of the relativity of motion to accelerated motion and it appeared that he had carried this belief over to his work on the general theory of relativity, in which a general equivalence of all states of motion is postulated. Einstein's original statement of the equivalence asserted the equivalence of uniform acceleration and homogeneous gravitational fields alone. Now it would appear natural to extend this equivalence to obtain between arbitrary states of acceleration and arbitrary gravitational fields. Then the final step in the above argument could be completed by a direct: "The principle of equivalence demands the (local) equivalence of all states of acceleration and all gravitational fields. Thus, in general, the tensor $g_{\mu\nu}$ must also characterise gravitational fields."

That such an extension had indeed been made is suggested by Einstein's accounts of the principle in a number of places. For example in his 1949 Autobiographical Notes he describes the inference that he made from the equality of inertial and gravitational mass:

> "In a gravitational field (of small spatial extension) things behave as they do in a space free of gravitation, if one introduces into it, in place of an 'inertial system', a frame of reference accelerated relative to the former." [27]

This appears to be meant to apply to arbitrary gravitational fields. But how consistently Einstein held this view in his later years is unclear, for another description of the principle from the same time clearly limits it to the case of uniform acceleration. [28]

---

27. A. Einstein, Autobiographical Notes (La Salle & Chicago, Illinois: Open Court, 1979), p.63.

28. See note 24.

Also, in his earlier years, particularly during the period when he was developing and refining the general theory of relativity, Einstein would often refer to the principle in terms suggesting an extension from uniform to arbitrary states of acceleration, but not explicitly stating it. Typical of these is his 1914 statement:

> "The whole theory stems from the conviction that all physical processes run exactly the same in a gravitational field, as the corresponding processes run without a gravitational field, in the case that one relates them to an appropriate accelerated (three dimensional) coordinate system ('Equivalence hypothesis')." 29

Einstein's discussion of the principle of equivalence does lead naturally to the idea that the principle should assert a general equivalence of arbitrary gravitational fields and states of acceleration. To the modern reader and perhaps to Einstein himself in his later years, this suggestion is reinforced by later accounts of the principle which explicitly extend its domain of application from homogeneous to arbitrary gravitational fields. However, in the crucial years including and following 1913, whilst Einstein was refining his general theory of relativity, we find from an examination of examples similar to those given above that he never explicitly made such an extension.

---

29. "Die ganze Theorie ist hervorgegangen aus der Überzeugung, dass alle physikalischen Vorgänge in einem Gravitationsfeld genau gleich ablaufen, wie die entsprechenden Vorgänge ohne Gravitationsfeld ablaufen, falls man sie auf ein passend beschleunigtes (dreidimensionales) Koordinatensystem bezieht ('Äquivalenzhypothese')."

    A. Einstein & M. Grossmann, "Kovarianzeigenschaften der Feldgleichhungen der auf die verallgemeinerten Relativitätstheorie gegründeten Gravitationstheorie", Zeitschrift für Mathematik und Physik, 63(1914), p.215.

On the contrary, whenever he did take the trouble to define exactly which gravitational fields and which states of acceleration were involved in the principle, he always limited himself to uniform acceleration and homogeneous gravitational fields. A typical example of this is in the 1916 statement of the principle quoted at the beginning of this section.

This feature of his understanding of the principle of equivalence is stated by him with great care and precision in a 1916 reply to a claim by Kottler that Einstein had given up the principle of equivalence in his general theory of relativity. (In the paragraph preceding this passage, Einstein tells us that the reference system "K" mentioned is one in which special relativity holds and covers a finite region of gravitation free spacetime. "Galilean system")

> "2. Principle of Equivalence. Going out of this boundary case of the special theory of relativity, one can ask oneself whether an observer, uniformly accelerated relative to K in the region considered, must understand his condition as accelerated, or whether there remains an understanding for him from the (approximately) known laws of nature, by which he can interpret his condition as 'rest'. Expressed more precisely: Do the laws of nature, known to a more certain approximation, allow us to consider as at rest a reference system K', which is accelerated uniformly with respect to K? Or somewhat more generally: Can the principle of relativity be extended also to reference systems, which are (uniformly) accelerated relative to oneanother? The answer runs: As far as we really know the laws of nature, nothing stops us from considering the system K' as at rest, if we assume the presence of a gravitational field (homogeneous in the first approximation) relative to K'; for all bodies fall with the same acceleration independent of their physical nature in a homogeneous gravitational field as well as with respect to our system K'. The assumption that one may treat K' as resting in all strictness without any laws of nature not being fulfilled with respect to K', I call the 'principle of equivalence'.
>
> "3. Gravitational fields not only kinematically conditioned. One can also invert the previous consideration. Let the system K', formed with the gravitational field considered above, be the original. Then one can introduce a new

reference system K, accelerated against K', with respect to which (isolated) masses (in the region considered) move uniformly in a straight line. But one may not go on and say: If K' is a reference system provided with an arbitrary gravitational field, then it is always possible to find a reference system K, in relation to which isolated bodies move uniformly in a straight line, i.e. in relation to which no gravitational field exists. The absurdity of such an assumption is quite obvious. If the gravitational field with respect to K', for example, is that of a stationary mass point, then this field certainly cannot be transformed away for the entire neighbourhood of the mass point even through any refined transformation trick. Therefore one may in no way assert that gravitational fields should be explained purely kinematically to some extent; a 'kinematic, not dynamic understanding of gravitation' is not possible. Therefore through mere transformations out of a Galilean system into another through acceleration transformations we do not to know arbitrary gravitational fields, but those of a quite special kind, which however must still satisfy the same laws as all other gravitational fields. This is only again another formulation of the principle of equivalence (special in its application to gravitation)." 30

---

30. "2. Äquivalenzprinzip. Ausgehend von diesem Grenzfall der speziellen Relativitätstheorie kann man sich fragen, ob ein in dem betrachteten Gebiete relativ zu K gleichförmig beschleunigter Beobachter seinen Zustand als beschleunigt auffassen muss, oder ob ihm nach dem (angenähert) bekannten Naturgesetzen eine Auffassung übrig bleibt, vermöge deren er seinen Zustand als 'Ruhe' deuten kann. Präziser ausgedrückt: Erlauben uns die in gewisser Annäherung bekannten Naturgesetze ein in bezug auf K gleichförmig beschleunigtes Bezugssystem K' als ruhend zu betrachten? Oder etwas allgemeiner: Lässt sich das Relativitätsprinzip auch auf relativ zueinander (gleichförmig) beschleunigte Bezugssysteme ausdehnen? Die Antwort lautet: Soweit wir die Naturgesetze wirklich kennen, hindert uns nicht daran, das System K' als ruhend zu betrachten, wenn wir relativ zu K' ein (in erster Annäherung homogenes) Schwerefeld als vorhanden annehmen; denn wie in einem homogenen Schwerefeld, so auch in bezug auf unser System K' fallen alle Körper unabhängig von ihrer Physikalischen Natur mit derselben Beschleunigung. Die Voraussetzung, dass man in aller Strenge K' als ruhend behandeln dürfe, ohne dass irgendein Naturgesetz in bezug auf K' nicht erfüllt wäre, nenne ich 'Äquivalenzprinzip'.

"3. Das Schwerefeld nicht nur kinematisch bedingt. Man kann die vorige Betrachtung auch umkehren. Sei das mit dem oben betrachteten Schwerefeld ausgestaltete System K' das ursprüngliche. Dann kann man ein neues, gegen K' beschleunigtes Bezugssystem K einführen, mit Bezug auf welches sich (isolierte) Massen (in dem betrachteten Gebiete) geradlinig gleichförmig bewegen. Aber man darf nicht weitergehen und sagen: Ist K' ein mit einem beliebigen Gravitationsfeld versehenes Bezugssystem, so ist stets ein Bezugssystem K auffindbar, in bezug auf welches sich isolierte Massen geradlinig gleichförmig bewegen, d.h. in bezug auf welches kein Gravitationsfeld existiert. Die Absurdität einer solcher Voraussetzung liegt auf der Hand. Ist das Gravitationsfeld in bezug auf K' zum Beispiel das eines ruhenden Massenpunktes, so lässt sich dieses Feld für die ganze Umgebung des Massenpunktes gewiss durch kein noch so feines Transformationskunststück hinwegtransformieren. Man darf also keineswegs behaupten, das Gravitationsfeld sei gewissermassen rein kinematisch zu erklären; eine 'kinematische, nicht dynamische Auffassung der Gravitation' ist nicht möglich. Durch blosse Transformation aus einem Galileischen System in ein anderes durch Beschleunigungstransformationen lernen wir also nicht beliebige Gravitationsfelder kennen, sondern solche ganz spezieller Art, welche aber doch denselben Gesetzen genügen müssen wie alle anderen Gravitationsfelder. Dies ist nur wieder eine andere Formulierung des Äquivalenzprinzips (speziell in seiner Anwendung auf die Gravitation)."

A. Einstein, "Über Friedrich Kottlers Abhandlung 'Über Einsteins Äquivalenzhypothese und die Gravitation'", Annalen der Physik, 51 (1916), pp.640-1. Einstein's emphasis.

Here Einstein is quite definite about the content of the principle of equivalence: it asserts the global equivalence of uniform acceleration and homogeneous gravitational fields. He also explains why he cannot entertain the extension of such a global equivalence to arbitrary gravitational fields. But no mention is made of such an extension with regard to the more limited case of a local equivalence, Such an extension has been entertained by many other relativists. [31]

From this, Einstein proceeds to a somewhat puzzling conclusion:

> "Therefore a gravitation theory violates the principle of equivalence, in the sense which I nderstand it, only then if the equations of gravitation are fulfilled in no reference system K', which moves non-uniformly relative to a Galilean reference system. It is evident that this objection cannot be raised against my theory with generally covariant equations; for here the equations are fulfilled with respect to each reference system. The demand of the general covariance of equations embraces that of the principle of equivalence as a quite special case." [32]

Now if the principle of equivalence is completely contained within the demand of general covariance, then it would seem that the principle is logically superfluous to the foundations of the general theory of

---

31. For example W. Pauli, Theory of Relativity (Oxford: Pergammon, 1958), p.145.

32. "Eine Gravitationstheorie verletzt also das Äquivalenzprinzip in dem Sinne, wie ich es verstehe, nur dann, wenn die Gleichungen der Gravitation in keinem Bezugssystme K' erfüllt sind, welches relativ zu einem galileischen Bezugssystem ungleichförmig bewegt ist. Dass dieser Vorwurf gegen meine Theorie mit allgemein kovarianten Gleichunger nicht erhoben werden kann, ist evident; denn hier sind die Gleichungen bezüglich eines jeden Bezugssystem erfüllt. Die Forderung der allgemeinen Kovarianz der Gleichungen umfasst die des Äquivalenzprinzips als ganz speziellen Fall."

Einstein, "Über Friedrich Kottlers...", p.641. Einstein's emphasis relativity. So, if this is the case, we might well ask what function the principle serves. Perhaps Einstein's answer can be found in the examination of a 1913 justification of the conclusion that the fundamental tensor accounts for both inertial and gravitational phenomena. In this justification, Einstein uses the principle of equivalence in the form described above, in which it asserts the equivalence of uniform acceleration and homogeneous gravitational fields.

> "From the usual theory of relativity an isolated material point moves uniformly in a straight line according to the equation
>
> $$\delta\left(\int ds\right) = 0$$
>
> where
>
> $$ds^2 = -dx^2 - dy^2 - dz^2 + c^2dt^2$$
>
> and c is the (constant) speed of light. Now the equivalence hypothesis admits the conclusion that a material point moves in a static gravitational field (of a special kind) according to the same equation, where however, c is a function of place and is determined through the gravitational potential. From this special case of the gravitational field one can reach in any case to a general one, by going over to moving coordinate systems through coordinate transformation..."

To this last sentence Einstein appended the crucial footnote:

> "Thereby we postulate that we attain an equally justified description of processes by relating them to an appropriately moving coordinate system; with this we hold fast to the basic thinking of the theory of relativity."

He continued:

> "...One recognises on this path that the only sufficiently extensive generalisation of the stated law of motion is in this: we assume the 'line element ds' has the form
>
> $$ds^2 = \sum_{ik} g_{ik}dx_i dx_k, \qquad (i,k=1,2,3,4)$$
>
> where the g_{ik} are functions of x_1, x_2, x_3 and x_4 and the first three coordinates characterise place, the last time, and the equation of motion again should have the form
>
> $$\delta\left(\int ds\right) = 0.$$
>
> If one considers that by this understanding in place of the usual line element
>
> $$ds^2 = \sum_i dx_i^2$$

of the original theory of relativity, the more general

$$ds^2 = \sum_{ik} g_{ik} dx_i dx_k$$

enters as the absolute invariant (scalar), then one recognises immediately how one attains a generalisation of relativity theory, which embraces gravitation on the foundation of the equivalence hypothesis."[33]

---

33. "Nach der gewöhnlichen Relativitätstheorie bewegt sich ein isolierter materieller Punkt geradlinig-gleichförmig gemäss der Gleichung

$$\delta\left(\int ds\right) = 0$$

wo

$$ds^2 = -dx^2 - dy^2 - dz^2 + c^2 dt^2$$

ist und c die (konstante) Lichtgeschwindigkeit bedeutet. Die Äquivalenzhypothese lässt nun die Folgerung zu, dass sich in einem statischen Schwerefeld (spezieller Art) ein materieller Punkt gemäss der nämlichen Gleichung bewegt, wobei aber c eine Funktion des Ortes ist und durch das Gravitationspotential bestimmt wird. Von diesem Spezialfall des Schwerefeldes kann man zu einem allgemeinen jedenfalls gelangen, indem man durch Koordinatentransformation auf bewegte Koordinatensysteme ubergeht.[1]) Man erkennt auf diesem

Wege, dass die einzige invarianten-theoretisch genügend umfassende Verallgemeinerung des angegebenen Bewegungsgesetzes darin besteht, dass wir das „Linienelement ds" in der Form

$$ds^2 = \sum_{ik} g_{ik} dx_i dx_k \qquad (i,k = 1,2,3,4)$$

voraussetzen, wo die g_{ik} Funktionen von x_1, x_2, x_3 und x_4 sind und die drei ersten Koordinaten den Ort, die letzte die Zeit charakterisieren und die Bewegungsgleichung wieder die Form

$$\delta\left(\int ds\right) = 0$$

haben soll.

Berücksichtigt man, dass bei dieser Auffassung an Stelle des gewöhnlichen Linienelementes

$$ds^2 = \sum_{i} dx_i^2$$

der ursprünglichen Relativitätstheorie das allgemeinere

$$ds^2 = \sum_{ik} g_{ik} dx_i dx_k$$

als absolute Invariante (Skalar) tritt, so erkennt man sofort, wie man zu einer Verallgemeinerung der Relativitätstheorie gelangt, welche auf der Grundlage der Äquivalenzhypothese die Gravitation mit umfasst."

1) Dabei postulieren wir, dass wir zu einer gleichberechtigten Beschreibung des Vorganges gelangen, indem wir ihn auf ein geeignet bewegtes Koordinatensystem beziehen; damit halten wir an dem Grundgedanken der Relativitätstheorie fest.

Here Einstein was not yet using the convention of raising and lowering contra and covariant indices, nor was he using the Einstein summation convention.

A. Einstein, "Physikalische Grundlagen einer Gravitationstheorie", <u>Naturforschende Gesellschaft, Vierteljahrsschrift (Zürich)</u>, <u>58</u> (1913), pp.285-6.

The focus of this passage is not so much the principle of equivalence as the extension of the principle of relativity. This latter requirement led to the need for a generally covariant formalism, in which all coordinate systems are equally admissible for the description of physical processes.

Einstein begins with a result which he says is drawn from the principle of equivalence, that the metric of a static gravitational field is of the Lorentzian form given, with the speed of light c varying as a function of place. In a later chapter, I will examine how this result follows from Einstein's invocation of the equivalence of uniform acceleration and homogeneous gravitational fields. We shall see that this result does not hold within the final form of the general theory of relativity and that this was a misconception for which Einstein would pay dearly.

In the context of this passage, however, this application of the principle of equivalence seems to serve only as an illustration of how the range of admissible coordinates can be extended, presumably to those which are uniformly accelerated, by the inclusion of gravitation into the theory. For Einstein then steps from this "special case" to a "general one", by introducing arbitrary coordinate transformations. Einstein makes it very clear that this latter extension is governed by the requirements of an extended principle of relativity.

Thus here Einstein seems to treat the principle of equivalence as providing an illustrative and suggestive "warm-up" exercise in the context of the problem extending the admissible range of coordinate systems for formulation of physical laws. It shows how this range of coordinate systems can be extended to the case of uniform acceleration by the appropriate inclusion of gravitation within the theory. But its function of extending this range is now completely usurped by an explicit invocation of the extended principle of relativity.

This would seem to explain what Einstein had in mind when he later wrote of the principle of equivalence as entirely contained within the requirement of general covariance. It also makes clear that Einstein has gone too far with this claim. For only one function of the principle of equivalence has been subsumed by the requirement of general covariance, that of justifying an extension of the range of admissible coordinate systems. What the requirement of general covariance cannot do is justify the conclusion that the behaviour of freely moving bodies, as seen from arbitrary coordinate systems, may be regarded equally well as directed by inertial or gravitational forces. That is that the fundamental tensor accounts for gravitational as well as inertial phenomena.

This latter conclusion requires an additional assumption about the nature of gravitation and inertia, an assumption which had been contained implicitly in the original versions of the principle of equivalence. It is that gravitation and inertia are essentially related phenomena and thus can be dealt with by the same theoretical machinery. This, of course, is the import of Einstein's 1918 statement of the principle of equivalence - that inertia and weight are the same in essence - and the conclusion that he drew from it - that both gravitation and inertia are to be accounted for by the fundamental tensor.

Thus the historical development of the principle of equivalence seems to run as follows. To begin with, the principle of equivalence bore the dual load of justifying a limited extension of admissible coordinate systems and also the identification of inertial and gravitational phenomena. With the advent of the general theory of relativity in its earliest form in 1913, this first function of the principle was subsumed by the more powerful extended principle of relativity and the principle of general covariance. This left the second function as the only contribution of the principle of equivalence to the foundations of the theory. At first Einstein seemed to overlook this contribution, although it always figured implicitly in his developments of the theory. By 1918, however, he had come to acknowledge its importance explicitly and he stated only this second aspect of the principle in his 1918 definition of it.

Thus a little later in this same 1918 paper, Einstein referred to the relationship between the principles of equivalence and general covariance in the following way:

> "Principle (b) [the principle of equivalence] constituted the starting point of the whole theory and first brought with it the postulation of principle (a); it can certainly not be given up as long as one wants to hold onto the basic thoughts of the theoretical system." 34

("Principle (a)" is called the "principle of relativity" in the paper, but from the definition given is actually the principle of general covariance.)

---

34. "Das Prinzip (b) hat den Ausgangspunkt der ganzen Theorie gebildet und erst die Aufstellung des Prinzipes (a) mit sich gebracht; es kann sicherlich nicht verlassen werden, solange man am Grundgedanken des theoretischen Systems festhalten will." Einstein, "Prinzipielles..." (1918), p.242.

To conclude this account of Einstein's use of the principle of equivalence in his general theory of relativity, we can confirm that it is actually this version of the principle which enables him to complete the argument which establishes that the fundamental tensor accounts for gravitational as well as inertial phenomena. This emerges somewhat indirectly in a 1914 development of this result. In the usual way, he arrives at the equation of motion (4.5) (his equation (2)) with the general metric (4.4b) (His equation (1a)) for the motion of a free mass point in gravitation free space, as seen from arbitrary coordinates. He then argues that these equations will also describe the motion of a free mass point under the action of a gravitational field:

> "This law of motion ((2), (1a)) is only derived immediately for the case that the point moves quite force free and therefore that no gravitational field acts on the point (judged from an appropriate reference system). But since according to experience the law of motion of a material point in the gravitational field does not depend on the material of the body, and since it is possible in any case to bring that law to a Hamiltonian form, thus this suggests that in general ((2), (1a)) are to be regarded as the law of motion of a point on which no forces other than gravitational forces act. This is the core of the 'equivalence hypothesis'." [35]

---

35. "Dies Bewegungsgesetz ((2),(1a)) ist zunächst nur für den Fall abgeleitet, dass sich der Punkt ganz kräftefrei bewegt, dass also auf den Punkt auch kein Gravitationsfeld einwirkt (von einem geeignet Bezugssystem aus beurteilt). Da aber erfahrungsgemäss das Bewegungsgesetz eines materiellen Punktes im Schwerefelde nicht vom Material des Körpers abhängt, und da es jedenfalls möglich sein dürfte jenes Gesetz auf die Hamiltonsche Form zu bringen, so liegt es nahe, ((2),(1a)) allgemein als das Bewegungsgesetz eines Punktes anzusehen, auf welchen keine anderen als Gravitationskräfte wirken. Dies ist der Kern der 'Äquivalenzhypothese'."

A. Einstein, "Prinzipielles zur verallgemeinerten Relativätstheorie und Gravitationstheorie", Physikalische Zeitschrift, 15 (1914), p.177.

Here we see that the crucial property of gravitational fields, which justifies Einstein's conclusion that they can be represented by the fundamental tensor, is that the motion of a free body in such a field is independent of the material (and, of course, the mass) of the body. This is also a characteristic property of inertial fields. Thus it seems quite appropriate here to use the same theoretical machinery - the fundamental tensor - to account for both gravitational and inertial phenomena, provided one is willing to assume a fundamental connection between the two phenomena. In other words, provided one is willing to assume the 1918 version of the principle of equivalence, which asserts that inertia and weight are the same in essence.

In this 1914 passage, Einstein tantalisingly refers to "this" as the "core of the 'equivalence hypothesis'". However, beyond this he does not make any clearer what he takes the hypothesis to be. The principle of equivalence is not defined elsewhere in the paper, for example. Indeed, he only gives it the explicit form of asserting the essential identity of inertia and weight in 1918.* But it is interesting to note that if we look at other examples of how Einstein establishes the significance of the fundamental tensor, this same crucial property of gravitational fields is sometimes used in the same way.[36]

---

* In the introduction to the "Entwurf..." paper, Einstein gives the familiar statement of the principle of equivalence as asserting the equivalence of uniform acceleration and homogeneous gravitational fields. He later refers to the principle as asserting the "physical equality of nature (physikalische Wesensgleichheit) of gravitational and inertial mass". Whether this is intended to be a reformulation of the older version of the principle for the new theory or just a summary of the older version is unclear. In any case, it does not reappear until 1918. Einstein, "Entwurf...", pp.225-6.

36. See, for example, Einstein, "The Foundation...", p.120.

It is interesting that, after the apparently satisfactory resolution of the question of the content and role of the principle of equivalence of the 1918 paper, Einstein reverted to his earlier statement of the principle of equivalence as requiring the equivalence of uniform acceleration and homogeneous gravitational fields. (See, for example, his 1921 Princeton lectures.) [37] Perhaps this can be explained by the fact that these later statements were part of essentially pedagogic and occasionally historical expositions. Einstein may have felt that, for these purposes, the earlier statement was to be preferred.

To summarise, we may extract the following picture of the shift in content and status of the principle of equivalence with the emergence in 1913 and later completion of the general theory of relativity.

From his earlier work on gravitation, Einstein took a principle of equivalence which asserted the equivalence of uniform acceleration and homogeneous gravitational fields. This equivalence had been the core of the scalar theory, being explicitly invoked throughout the theory to determine the effects of gravitation on many processes. It had led to the extension of the range of frames of reference admissible for the formulation of physical laws to include those in uniform acceleration and this had been achieved by the expedient of including gravitational fields in the theory of space and time.

With the emergence of the general theory of relativity, the principle of equivalence came to contribute only one essential but preliminary result: that gravitational and inertial phenomena are

---

37. A. Einstein, The Meaning of Relativity (London: Chapman & Hall, 1976) pp.55-7.

fundamentally connected and are both to be accounted for by the fundamental tensor. Einstein was slow to give explicit recognition to this shfit in the content and import of the principle. At first he focussed on the principle in its role of extending the range of frames of reference admissible for the formulation of physical laws. He recognised that this function had been completely usurped by the extended principle of relativity and the principle of general covariance in the new theory. So he temporarily described the principle of equivalence as being completely contained within the principle of general covariance. By 1918, he recognised that the principle also contained a quite separate hypothesis on the nature of gravitation and inertia, that inertia and weight are the same in essence. This came to be his new statement of the content of the principle and he saw its expression in the theory in the fact that both inertial and gravitational phenomena are represented by the fundamental tensor.

4.6.3 The Independence of the Acceleration of Free Fall

In 1907, in a remarkable display of foresight and intuition, Einstein recognised that gravitational fields were distinguished by a special property, whose full significance had not yet been revealed by then current physical theories. In his 1911 Annalen der Physik paper he described this property:

> "This experience, of the equal falling of all bodies in the gravitational field, is one of the most universal which the observation of nature has yielded; but in spite of that the law has not found any place in the foundations of our edifice of the physical universe." [38]

---

38. Einstein, "On the Influence...", p.100.

In particular we have seen that Einstein insisted that the downward acceleration of a freely falling body was independent of its sideways velocity.

This belief was central to Einstein's early work on gravitation in relativity theory. It had justified his early conclusion that no special relativistic gravitation theory was possible and had enabled the postulation of the concept central to this early work, the equivalence of acceleration and gravitation.

In this section we shall see that, following the emergence of the general framework of the general theory of relativity in 1913, Einstein carried through a subtle but significiant reappraisal of the nature of this crucial property of gravitational fields.

To begin with, we find Einstein introducing his 1913 "Entwurf..." paper with a statement remarkably similar to the one quoted above. However, the fundamental property of gravitation which has not yet "found a place in the foundations of our edifice of the physical universe" is no longer the "equal falling of all bodies in the gravitational field":

> "The theory presented in the following springs out of the conviction that the proportionality between the inertial and gravitational mass of bodies should be an exactly valid natural law, which must now find an expression in the foundations of theoretical physics." [39]

The introduction continued with a brief mention of his earlier work and the "equivalence hypothesis". It concluded with a careful description of the Eötvös experiment. He concluded that this experi-

---

39. "Die im folgenden dargelegten Theorie ist aus der Überzeugung hervorgegangen, dass die Proportionalität zwischen der trägen und der schweren Masse der Körper ein exakt gültiges Naturgesetz sei das bereits in dem Fundamente der theoretischen Physik einen Ausdruck finden müsse." Einstein, "Entwurf...", p.225.

ment, in conjunction with the fact that its results applied to matter which had undergone radioactive decay, showed the exact equality of inertial and gravitational mass.

Superficially, the shift which we have seen here is innocuous. The two properties of gravitation concerned here, the uniqueness of the acceleration of free fall and the equality of inertial and gravitational mass, are often taken to be equivalent. Indeed an explanation for Einstein's shift comes readily to hand. Prior to 1913, we might suggest, he did not know of the Eötvös experiment. Somehow he came to know of it by the time of the "Entwurf..." paper and proceeded to cite it and its conclusion, the equality of inertial and gravitational mass, as the most accurate empirical result available on gravitation. This view is supported by Einstein's later recollections of the failure of his 1907 special relativistic gravitation theory:

> "This did not fit in with the old experimental fact that all bodies have the same acceleration in a gravitational field. This law, which may also be formulated as the law of the equality of inertial and gravitational mass, was now brought home to me in all its significiance...I had no serious doubts about its strict validity even without knowing the results of the admirable experiments of Eötvös, which - if my memory is right - I only came to know later." [40]

However, whatever the reason for Einstein's shift on this question, it is by no means innocuous. It is both significant and appropriate to his new theory. Two possible properties of gravitational fields have been delineated:

(a) that the downward acceleration of a freely falling body is independent of the nature of the body and its sideways velocity, and

(b) the equality of inertial and gravitational mass.

---

40. Einstein, "Notes on the Origin...", p.287.

These two properties are not equivalent. (a) does entail (b). Thus it was possible for Einstein to generate (b) as a theorem in his gravitation theories prior to 1913, which were based on the assumption of property (a). (b), however, does not entail (a), without further assumption. Newtonian gravitation theory provides exactly these necessary assumptions required for (b) to entail (a). Einstein's final general theory of relativity, however, predicts that gravitational fields have property (b), but not (a).

To demonstrate this, we need first recognise that the "equality of inertial and gravitational mass" is not a phenomenological requirement, as is property (a). Rather it is a postulate of Newtonian gravitation theory and one which is expressed in the language of that theory's theoretical terms. Fortunately this turns out to present no problems to us here. For in his 1913 accounts of the Eötvös experiment, Einstein made quite clear what he understood the equality of inertial and gravitational mass to mean in phenomenological terms.[41] It required

(b') that the magnitude and direction of the acceleration produced by the combination of inertial and gravitational forces acting on an otherwise free body be independent of the nature of the body.

It is clear that (a) entails (b'), for (b') is effectively the same as (a) except for the additional requirement in (a) that downward acceleration be independent of sideways velocity. Similarly (b') alone cannot entail (a).

---

41. See Einstein, "Entwurf...", pp.225-6 and Einstein, "Zum gegenwärtigen stande...", p.1251.

Einstein's final general theory of relativity is consistent with (b'), for in it the world lines of freely falling bodies are independent of the nature of the bodies. However, in this theory, the downward acceleration of a freely falling body is not independent of its sideways velocity. Consider, for example, the weak field which acts in the x-direction alone and thus has a metric

$$ds^2 = (c^2 + 2\varphi)\,dt^2 - \left(1 - \frac{2\varphi}{c^2}\right)(dx^2 + dy^2 + dz^2) \quad (4.6a)$$

where φ is a function of x alone A calculation of the geodesics of this field (see Appendix B) shows that the acceleration of fall of a free body varies with its sideways velocity.

Whether Einstein was aware of this implication of his theory in 1913 is not clear. He did associate the equality of inertial and gravitational mass with the indepedence of acceleration of bodies in free fall. So, for example, in his 1913 Vienna address he wrote of the postulate of the equality of the inertial and gravitational mass of a closed system in the following way:

> "This [postulate] is directly supported by the fact of experience that all bodies in a gravitational field fall with equal acceleration...Here let it be observed that the equality (proportionality) of gravitational and inertial mass was proven with great accuracy through an investigation by Eötvös of the greatest importance for us; Eötvös proved this proportionality by showing experimentally that the resultant of the weight and the centrifugal force, originating in the rotation of the earth, is indepdnent of the nature of the material (relative difference of both masses $< 10^{-7}$)."[42]

---

42. "Dasselbe stützt sich zunächst auf die Erfahrungstatsache, dass alle Körper im Schwerefelde mit gleicher Beschleunigung fallen... Hier sei bemerkt, dass die Gleichheit (Proportionalität) der schweren und trägen Masse durch eine für uns höchst wichtige Untersuchung von Eötvös mit grosser Genauigkeit erwiesen wurde; Eötvös wies diese Proportionalität nach, indem er experimentell zeigte, dass die Resultierende der Schwere und der von der Drehung der Erde herrühenden Zentrifugalkraft von der Natur des Materials unabhängig ist (relativer Unterschied beider Massen $< 10^{-7}$). Einstein, "Zum gegenwärtigen Stande...", p.1251

What is not made clear, however, is whether this equality of acceleration extends to bodies with different sideways velocities.

Now it may well have been that Einstein saw no point in distinguishing the two properties of gravitation delineated here since his 1913 "Entwurf..." theory predicted that gravitational fields had both properties. We shall see later that within this theory a static gravitational field which acts in the x direction only has the metric

$$ds^2 = (c^2 + 2\varphi)\,dt^2 - dx^2 - dy^2 - dz^2 \tag{4.6b}$$

where φ is a function of x only. A calculation of the geodesics of this field (Appendix B) shows that in this case the acceleration of fall of a free body is independent of its sideways velocity.

However, to say that he thought the distinction unimportant is not to say that he did not recognise it. We can be sure that by 1913 he had come to recognise that the equality of inertial and gravitational mass does not entail the independence of acceleration of fall from sideways velocity. This emerges from a study of Einstein's involvement with the development of Nordström's theory of gravitation in the period in question.

The emergence and development of Nordström's theory of gravitation is of special interest to us here. For Nordström continued with the task of constructing a scalar, special relativistic theory of gravitation from the point where Einstein had left it in 1907. Later I devote an entire chapter, Chapter 8, to an account of this episode and its significance. I refer the reader forward to this chapter for further details on the material which now follows.

We shall see that in 1912 Nordström proposed a scalar theory of gravitation which was essentially identical to the one outlined in Chapter 3 above. Einstein immediately informed Nordström that he had already considered a similar possibility but had rejected it because he believed that it could not possibly correspond with reality. Specifically he showed that, within Nordström's theory, a rotating body is accelerated less by a gravitational field than a non-rotating one. Clearly this follows from the result that the downward acceleration of each of the elements of the rotating body would be decreased by its sideways velocity, in accord with equation (3.20a).

In the following year Einstein returned to the question of a scalar theory of gravitation in the concluding section of his part of the "Entwurf..." paper. He no longer raised the above objection. Rather he argued that a scalar theory of gravitation could be brought into accord with the equality of inertial and gravitational mass if the source density term is taken to be the trace of the stress-energy tensor. With this choice, it followed within a scalar theory that the gravitational mass of a closed system in steady state, such as a confined gas or a uniformly rotating body, was given by the system's total energy and that its downwards acceleration was independent of its internal motions.

Einstein also made a new objection to the possibility of a scalar gravitation theory. Nordström was able to present a modified version of his theory, which used Einstein's suggestion for a source density term and which, was not open to Einstein's new objection. Einstein was also satisfied with the new version of the theory. In his 1913 Vienna address, later in that year, he presented an improved formulation and derivation of the theory. He concluded by noting that the theory met all known empirical requirements:

> "In conclusion we can say that Nordström's scalar theory, which retains the postulate of the constancy of the speed of light, complies with all conditions which can be placed on a theory of gravitation with the current state of experience..." [43]

However, Einstein did have an epistemological objection to the theory. In the theory the inertial mass of a body is influenced, but not caused by the other bodies of the universe. We shall see later that he used this as the point of departure for his general theory of relativity.

At the beginning of this 1913 address, Einstein listed four conditions which relativistic gravitation theories should seek to satisfy. The second required the "Equality of inertial and gravitational mass of closed systems". [44] The inclusion of the requirement that the system be closed was probably inspired by his work on the behaviour of the trace of the stress-energy tensor as a source density term in scalar theories. Later in the paper he went to some trouble to show that Nordström's theory was in accord with this postulate. He dealt with active and passive gravitational mass separately. He showed that both the gravitational force on a closed system and the

---

43. "Zusammenfassend können wir sagen, dass die Nordströmsche skalartheorie, welche an dem Postulat der Konstanz der Lichtgeschwindigeit festhält, allen Bedingungen entspricht, die an eine Theorie der Gravitation beim heutigen Stande der Erfahrung gestellt werden können...", ibid., p.1254.

44. "Gleichheit der trägen und der schweren Masse abgeschlossener systeme", ibid., p.1250.

number of gravitational lines of force sent out by such a system are proportional to its total energy. [45]

The use of Einstein's suggestion for a source term enabled Nordström's modified theory to predict that the acceleration of a closed system would be independent of its internal motions, provided their configuration was stationary. However, the theory still predicted that any sideways motion of the system as a whole would still affect its downward acceleration, according to a relation very close to equation (3.20a). Furthermore, we can be sure that Einstein was aware of this result. For in a 1913 paper, Nordström noted that Einstein had communicated an extended version of this result to him. (See Section 8.3)

This last conclusion is of great importance, for it finally confirms that a significant shift had occurred in Einstein's understanding of the basic empirical facts of gravitation. Einstein had begun his work on gravitation with the assumption that the downward acceleration of a freely falling body is independent of the nature of the body and, of great importance, independent of the body's sideways velocity. We have seen that this latter assumption lay at the heart of his 1907 conclusion that a special relativistic gravitation theory was not possible. It was also essential to his postulated equivalence of acceleration and gravitation.

In 1913 Einstein no longer regarded this independence of the downward acceleration of a freely falling body from its sideways velocity as a basic empirical fact of gravitation, although he may well have recognised it as a theorem of the 1913 "Entwurf..." theory.

---

45. Ibid., pp.1253-4.

The basic empirical fact of gravitation was now the weaker requirement of the equality of inertial and gravitational mass. He acknowledged that Nordström's theory, which satisfied the requirements of the equality of inertial and gravitational mass, "...complied with all known conditions which can be placed on a theory of gravitation with the current state of experience...". Yet this theory, as he was aware, predicted that the sideways velocity of a freely falling body would influence its downward acceleration.

4.7 Conclusion

In 1907 Einstein began work on a new theory of space, time and gravitation,which he hoped would ultimately lead to an extension of the principle of relativity to accelerated motion. At its heart lay Einstein's recognition of a fundamental property of gravitation, which had hitherto remained uninterpreted: the motion of a freely falling body is independent of its nature. On the strength of this, he had postulated the equivalence of acceleration and gravitation and the repeated invocation of this equivalence came to be central to the application of the theory. Einstein believed that the theory which resulted from this equivalence had at least realised an extension of the principle of relativity which encompassed uniform acceleration. He had had some success in extending the covariance of physical laws to include the case of uniform acceleration and static gravitational fields.

By 1912, the theory had been developed to an advanced and sophisticated level. However he was no closer to a general extension of the principle of relativity than he had been with the theory's earliest

1907 version. Later in 1912 and 1913 the major breakthrough came. His 1912 gravitation theory promised to provide an account of the origin of inertia as lying in an interaction between the masses of the universe. By 1913 he had published his "Entwurf..." paper, which contained an outline of the basic framework of his general theory of relativity.

However, with this breakthrough, Einstein came to shift and weaken the content and import of the principle of equivalence, as it had come to be called. This process was somewhat hidden by Einstein's discussions of the foundations of his new theory. Actually it had already begun in 1912. Einstein had found that his earlier theory did not even allow him to retain the equivalence of uniform acceleration and homogeneous gravitational fields as a global principle. It had to be weakened to a local one.

With the emergence of the "Entwurf..." theory, one of the major functions of the principle of equivalence, that of extending the range of frames of reference admissible for the formulation of physical laws, came to be usurped by an explicit invocation of the more powerful extended principle of relativity and the principle of general covariance. At first Einstein was uncertain about the independent contribution of the principle to the theory. He even suggested that the principle was completely contained within the principle of general covariance.

By 1918 Einstein recognised the independent contribution of the principle of equivalence to the theory. As well as governing an extension of the range of admissible frames of reference for the formulation of physical laws, the principle also contained an essential hypothesis about the nature of gravitation. It was that gravitational and inertial phenomena are essentially related and can be accounted for by the same theoretical machinery. This latter result was the only independent conclusion contributed by the principle to the theory. So Einstein now took this as the statement and import of the principle.

To be specific, Einstein concluded that both inertial and gravitational phenomena were to be accounted for by the fundamental tensor. We might regard this as the ultimate expression in the theory of the special property of gravitation, the uniqueness of free fall, which Einstein had first noticed in 1907 and had become the point of departure of the whole enterprise. A more modern view would take the expression of this special property to be that gravitational and inertial phenomena can be represented by an affine connection. Presumably Einstein would have agreed had the significance of the affine connection been recognised at the time.

Beneath this lay another shifting and weakening, this time of what Einstein took to be the basic empirical facts of gravitation. In 1907 he took one of these to be that the downward acceleration of a freely falling body was independent of the nature of the body and, in particular, of its sideways velocity. By 1913 Einstein had very quietly come to drop this latter requirement. He now took the basic empirical fact of gravitation to be the equality of inertial and gravitational mass, which entailed the independence of the downward acceleration of a freely falling body from its nature, but not its sideways velocity.

We can now recognise this as a serious blow to Einstein's earliest conclusions on gravitation. It will be recalled that the empirical basis of his rejection of the possibility of a special relativistic gravitation theory lay in his belief that the downward acceleration of a freely falling body is unaffected by it sideways velocity. Similarly, this independence is essential to the equivalence of acceleration and gravitation. If it does not stand, then a co-moving observer can distinguish free fall in a gravitational field from the absence of a gravitational field by watching the behaviour of bodies with "sideways" motion.

CHAPTER 5

THE PRINCIPLE OF EQUIVALENCE II

5. THE PRINCIPLE OF EQUIVALENCE II

In 1907 Einstein set forth on his journey towards a relativistic gravitation theory and a general theory of relativity armed with a bold postulate, the equivalence or indistinguishability of uniform acceleration and homogeneous gravitational fields. In his 1907 Jahrbuch... article, he made clear how this earliest version of the principle of equivalence was going to be exploited by him:

> "The heuristic value of this assumption lies therein that it makes possible the replacement of a homogeneous gravitational field by a uniformly accelerated reference system, the latter case being amenable to theoretical treatment to a certain degree." [1]

In the years following, we have seen how Einstein repeatedly used the principle in this way, as a bridge which would enable him to pass from inertial fields, whose structure was accessible to him through simple kinematic arguments, to the more obscure structure of gravitational fields.

In spite of Einstein's sustained faith in the principle, its status in the theory which finally emerged, the general theory of relativity, has been called into question and its importance claimed to be only historical and pedagogical. North, for example, writes:

> "The principle is of historical importance, although it need be given no place in a non-historical exposition of the general theory of relativity. Its greatest value seems to lie in its convincing hardened workers in the older tradition of the conceptual simplicity of the new theory." [2]

---

1. A. Einstein, "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen", Jahrbuch der Radioaktivität und Elektronik, 4(1907), p.454; translation from H.M. Schwartz, "Einstein's Comprehensive 1907 Essay on Relativity", American Journal of Physics, 45(1977), p.899.

2. J.D. North, The Measure of the Universe (Oxford: Clarendon, 1965), p.54.

In this chapter, I seek to determine the role of Einstein's postulation of the equivalence of acceleration and gravitation in the development of his general theory of relativity. In particular, from the perspective of the general theory of relativity, I examine the prospects of and difficulties facing Einstein in his attempts to construct a relativistic gravitation theory on the basis of this equivalence. In the following section I begin by examining the status of this equivalence in the final general theory of relativity. That is, I examine the extent to which the theory will allow acceleration to stand for gravitation.

5.1 The Equivalence of Acceleration and Gravitation

The determination of the extent to which an equivalence of acceleration and gravitation can be asserted within the general theory of relativity is hampered by the proliferation of different "principles of equivalence" which have appeared in various expositions of the theory and which assert an equivalence of acceleration and gravitation to one degree or another. Typical of the most traditional form of the principle is Pauli's 1921 version:

> "For every infinitely small region (i.e. a world region which is so small that the space- and time-variations of gravity can be neglected) there always exists a coordinate system $K_o(X_1,X_2,X_3,X_4)$ in which gravitation has no influence either on the motion of particles or on any other physical processes. In short, in an infinitely small world region every gravitational field can be transformed away." [3]

Other eminent relativists have used essentially the same statement of the principle and, with Pauli, have related it to the quadratic

3. W. Pauli, Theory of Relativity, 2nd.ed.(1921; Oxford : Pergammon, 1958), p.145.

form of the metric.[4] Others, however, present weaker or more ambiguous statements as their "principle of equivalence". Bergmann, for example, gives the principle as the requirement of the equality of gravitational and inertial mass and locates it in Newtonian theory.[5] Landau and Lifshitz follow a similar line of argument.[6]

In recent years, some effort has been put into distinguishing various forms of the traditional version of the principle, as stated by Pauli above. This stems from the work of Dicke and his colleagues.[7] "Weak", "Einstein" and "strong" forms of the principle have been distinguished and carefully defined.[8]

In general, these principles of equivalence all arise from attempts to state the significance to gravitation theory of one of the most characteristic properties of gravitational fields. This

---

4. For example R.C. Tolman, Relativity, Thermodynamics and Cosmology (1934; Oxford: Clarendon, 1958), p.175; H. Reichenbach, The Philosophy of Space and Time (1927; New York: Dover, 1958), pp.222-32, 251-2; M. Born, Einstein's Theory of Relativity (1920; New York: Dover, 1962), pp.312-7, 335-8; H.P. Robertson and T.W. Noonan, Relativity and Cosmology (Philadelphia: Saunders, 1968), p.171.

5. P.G. Bergmann, Introduction to the Theory of Relativity (1942; New York: Dover, 1976), p.153.

6. L. Landau and E.M. Lifshitz, The Classical Theory of Fields (Reading, Mass.: Addison-Wesley, 1959), pp.244-5. See also C. Møller, The Theory of Relativity (Oxford: Clarendon, 1952), pp.220-2.

7. R.H. Dicke, The Theoretical Significance of Experimental Relativity (New York: Gordon & Breach, 1964); P.G. Roll, R. Krotkov and R.H. Dicke, "The Equivalence of Inertial and Passive Gravitational Mass", Annals of Physics, 26 (1964), pp.442-517.

8. K.S. Thorne, D.L. Lee and A.P. Lightman, "Foundations for a Theory of Gravitation Theories", Physical Review D, 7 (1973), pp.3570-2.

property, often described as the equality of inertial and gravitational mass, manifests itself in the fact that the motion of a freely falling body in a gravitational field is in large measure independent of the body's internal structure and composition. It will be recalled that Einstein had been struck by this feature of gravitation in 1907 and it had led him to postulate the first form of the principle of equivalence.

A careful statement of this property comprises the "weak equivalence principle". Misner, Thorne and Wheeler give it the less misleading title of the "uniqueness of free fall".[9] This property enables us to regard spacetime as filled with a unique set of lines, the trajectories of bodies in free fall. The independence of these lines from the nature of the bodies falling enables us to identify them with the set of geodesics of a metrical spacetime. In 1918 Einstein recognised this latter statement as the direct theoretical embodiment of the special property of gravitation in his theory, as will be recalled from the previous chapter.

This interpretation of the significance of the uniqueness of free fall is fairly straightforward. Difficulties arise, however, when the attempt is made to draw out further implications of the uniqueness of free fall in some kind of an equivalence of acceleration and gravitation, where this latter term is understood to mean an experimental indistinguishability of the effects of acceleration and gravitation.

To begin with, the general theory of relativity provides an unambiguous way of distinguishing gravitational and pure inertial

---

9. C.W. Misner, K.S. Thorne and J.A. Wheeler, Gravitation (San Franscisco: Freeman, 1970), p.1050.

fields. The former are characterised by the presence of tidal gravitational forces and the latter by their absence. On the formal level, these tidal forces are described by the components of the fourth rank Riemann curvature tensor. If all its components vanish, then there are no tidal forces; spacetime is flat at the event in question; any "gravitational" forces present are only apparent and caused by the acceleration of the observer against a gravitation free space. Correspondingly, if the components of this curvature tensor do not vanish, then the field is at least in part a true gravitational field and spacetime at the event in question is curved.

The crucial feature of this criterion is that it is coordinate independent. The curvature of spacetime is independent of the coordinate system from which it is viewed. If there are no tidal forces and the Riemann curvature tensor vanishes in one coordinate systen, then it will do so in all coordinate systems.

This clearly rules out the possibility of an unrestricted equivalence of acceleration and gravitation in the general theory of relativity. Because of this a number of prominent relativists have denied a principle of equivalence any place in the logical structure of the theory. One of the more celebrated statements of this point of view is due to Synge:

> "...I have never been able to understand this Principle [of Equivalence] . Does it mean that the signature of the space-time metric is +2 (or -2 if you prefer the other convention)? If so, it is important, but hardly a Principle. Does it mean that the effects of a gravitational field are indistinguishable from the effects of an observer's acceleration? If so, it is false. In Einstein's theory, either there is a gravitational field or there is none according as the

> Riemann tensor does not or does vanish. This is an absolute property; it has nothing to with an observer's world-line. Space-time is either flat or curved..."[10]

Other relativists, such as Fock and Bondi, have taken a similar view.[11] Their predecessors can be traced back to as early as Eddington's 1924 exposition of relativity theory.[12] Eddington concedes the importance of the principle to the history of the theory, but, taking a position similar to Synge, argues that it no longer has a place in the logical foundations of the final theory. This point has been taken up in the modern literature.[13]

But, whilst the general theory of relativity will not allow a thoroughgoing equivalence of acceleration and gravitation, it does allow a restricted equivalence. This equivalence is generally limited to non-gravitational phenomena and regions of spacetime of small extent. In this area, its clearest expression is in the "Einstein" form of the equivalence principle. This asserts the undetectibility of the acceleration of free fall to any local non-gravitational experiment in a freely falling frame of reference. This implies that non-gravitational laws of nature revert to their

---

10. J.L. Synge, Relativity: The General Theory (Amsterdam: North Holland, 1960), p.ix.

11. V. Fock, The Theory of Space, Time and Gravitation (London: Pergammon, 1959), pp.206-11; H. Bondi, "Relativity Theory and Gravitation", in Einstein: A Centenary Volume, ed. A.P. French (Cambridge, Mass.: Harvard University Press, 1979), pp.113-29.

12. A.S. Eddington, The Mathematical Theory of Relativity (1924; Cambridge: Cambridge University Press, 1963), pp.39-41.

13. M. Sachs, "On the Logical Status of Equivalence Principles in General Relativity Theory", British Journal for the Philosophy of Science, 27 (1976), pp.225-9.

special relativistic forms in such a frame of reference. This result has been found to be of great use in enabling the construction of general relativistic forms of non-gravitational laws from their special relativistic forms by the "comma-goes-to-semi-colon rule".[14]

As well as allowing a local equivalence of acceleration and gravitation for non-gravitational phenomena, the general theory of relativity has also been understood to allow this local equivalence to extend to gravitational phenomena as well, in the more traditional versions of the principle. This stems from the theory's prediction that an acceleration of the appropriate magnitude and direction can mimic or conceal the effects of a given gravitational field at a point in spacetime provided that the metric tensor and its first derivatives only are considered. So, under this condition, the structure of spacetime reverts to that of special relativity at a point in a freely falling reference frame.

However, this effect does not amount to a local equivalence of acceleration and gravitation for gravitational phenomena. At every non-singular point in spacetime there still remains an unambiguous arbiter of the presence or absence of a true gravitational field, the Riemann curvature tensor, which is constructed out of the metric tensor and its first <u>and</u> second derivatives. Correspondingly, on the experimental level, tidal gravitational forces, which are associated with the Riemann curvature tensor and the second derivatives of the metric tensor, can still be used to distinguish inertial and gravitational fields on the a local level.

---

14. Misner <u>et al</u>., pp.385-7.

Ohanian has analysed the behaviour of various effects attributable to tidal gravitational forces as the system in question is made arbitrarily small and has found that they do not vanish. For example, the shape of a drop of liquid, bulged by tidal forces, remains even as the drop is made arbitrarily small, excluding surface tension effects. Also gravity gradiometers of arbitrarily small size can still detect tidal components in a gravitational field, allowing for construction difficulties and quantum effects. [15]

That the effects of acceleration and gravitation are not equivalent as far as gravitational phenomena are concerned manifests itself in a number of qualitatively interesting effects. In the last chapter we have seen that the downward acceleration of a test body varies, in general, with its sideways velocity. Further, it has been shown that the trajectory of a spinning particle in free fall deviates from the usual geodesics [16] and, if one considers the gravitational or electromagnetic field energy of the test body, then bodies of larger mass or charge fall slower. [17] Finally, since accelerated charges radiate, we would expect charges stationary in

---

15. H.C. Ohanian, "What is the Principle of Equivalence?" American Journal of Physics, 45 (1977), pp.903-9.

16. A. Papapetrou, "Spinning Test Particle in General Relativity I", Proceedings of the Royal Society, London, A209 (1951), pp.248-58; E. Corinaldesi and A. Papapetrou, "Spinning Test Particles in General Relativity II", Proceedings of the Royal Society, London, A209 (1951), pp.259-68.

17. C.H. McGruder III, "Field Energies and Principles of Equivalence", Nature, 272 (1978), pp.806-7.

a gravitational field to radiate as well, if the effects of acceleration and gravitation are really identical. This, of course, does not happen.[18]

5.2 The Kinematics of Acceleration

Before turning to a reexamination of Einstein's early work on gravitation, I present a number of results concerned with the kinematics of acceleration, which will be of use later.

Since the earliest days of the general theory of relativity, transformation equations, which relate inertial coordinate systems to those in states of constant proper acceleration, have been known.[19] If K(x,y,z,t) accelerates with constant proper acceleration g in the x' direction of the inertial coordinate system K'(x',y',z',t'), then K is related to K' according to

$$\begin{aligned} x' &= x \cosh \frac{gt}{c_L} + \frac{c_L^2}{g}\left(\cosh \frac{gt}{c_L} - 1\right) \\ y' &= y \qquad z' = z \\ t' &= \frac{c_L}{g} \sinh \frac{gt}{c_L} + \frac{x}{c_L} \sinh \frac{gt}{c_L} \end{aligned} \qquad (5.1a)$$

where the origin and corresponding coordinate axes of K and K' coincide when t=t'=0 and c_L is used to denote the speed of light in gravitation free Lorentzian space.*

---

18. H. Bondi and T. Gold, "The Field of a Uniformly Accelerated Charge, with Special Reference to the Problem of Gravitational Acceleration", Proceedings of the Royal Society, London, A229 (1955), pp.416-24.

19. F. Kottler, "Fallende Bezugssyteme vom Standpunkte des Relativitätsprinzips", Annalen der Physik, 45 (1914), pp.481-516; F. Kottler, "Über Einsteins Äquivalenzhypothese und die Gravitation", Annalen der Physik, 50 (1916), pp.955-72. See also Fock, pp.206-11.

* c_L is used to represent this quantity, in this chapter only, to avoid confusion with the symbol c, which will be used to denote the variable speed of light in Einstein's scalar theory of gravitation.

The metric in K' is

$$ds^2 = c_L^2 dt^2 - dx^2 - dy^2 - dz^2 \tag{5.2a}$$

Substituting directly from the transformation (5.1a), the metric in the associated system K is found to be

$$ds^2 = \left(c_L + \frac{gx}{c_L}\right)^2 dt^2 - dx^2 - dy^2 - dz^2 \tag{5.2b}$$

This can be rewritten in the form

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2$$

where

$$c = \left(c_L + \frac{gx}{c_L}\right) \tag{5.3a}$$

This defines the properties of the inertial field produced by constant proper acceleration. From the form of the metric (5.2b), we can see that the behaviour of measuring rods shows no effect due to the presence of the field; the spatial geometry revealed by them remains Euclidean. However, the speed of clocks will vary with position in the x-direction of the field and the speed of a light pulse, which follows a null geodesic, will be given by c in equation (5.3a).

If the transformation is rewritten for small t, ignoring terms in t^3 and higher, it takes the form

$$\begin{aligned} x' &= x + \tfrac{1}{2} g \frac{c}{c_L} t^2 \\ y' &= y \qquad z' = z \\ t' &= \frac{c}{c_L} t \end{aligned} \tag{5.1b}$$

The important relation describing the variation of the period of a clock, dT, at rest in K, from the coordinate time differential dt follows directly from the metric (5.2b):

$$dT = \frac{ds}{c_L} = \left(1 + \frac{gx}{c_L^2}\right) dt \tag{5.4a}$$

Provided t is small, the proper time T, indicated by clocks at rest in K, will not differ significantly from t' and, to the order of approximation of (5.1b), equation (5.4a) will reduce to

$$T = \left(1 + \frac{gx}{c_L^2}\right) t \tag{5.4b}$$

which also follows directly from (5.1b) by substitution with (5.3a).

5.3 The Basis of Einstein's Scalar Theories of Gravitation

The postulated equivalence of uniform acceleration and homogeneous gravitational fields was the central concept of the scalar theories of static gravitational fields which Einstein developed in the 1907-1912 period. He based these theories on the belief that this equivalence could act as a bridge which would enable him to pass from the accessible structure of inertial fields to the otherwise far less accessible structure of gravitational fields.

In Figure (5.1), I schematically outline the argument which is fundamental to these theories, making use of the results of the previous chapter. The argument runs as follows:

Step 1 From the kinematics of acceleration in a Lorentzian spacetime, we conclude that spacetime, in the presence of an inertial field produced by uniform acceleration, will have a structure that is described by the metric (5.2b). That is, the readings of clocks at rest in the field will vary from coordinate time according to equations (5.4a) and (5.4b); the speed of light, measured with coordinate times and distances, will be isotropic but will vary linearly with distance in the direction of the field, according to equation (5.3a); and the behaviour of three dimensional space, as given by the behaviour of both rods and coordinate distances, will remain Euclidean.

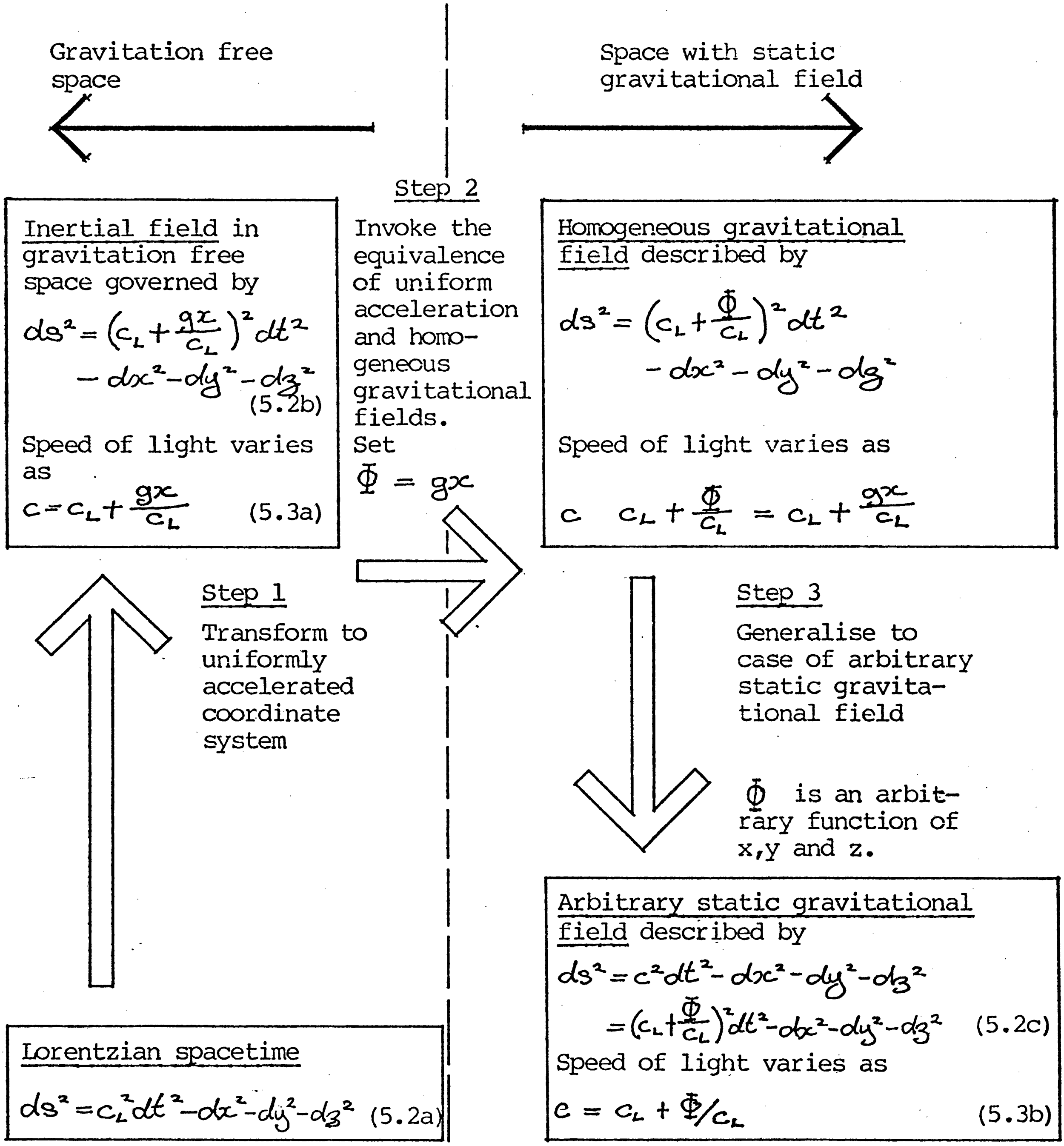


Figure 5.1 Foundation of Einstein's Scalar Theory of Static Gravitational Fields

Step 2 We invoke the equivalence of uniform acceleration and homogeneous gravitational fields and conclude that spacetime, in the presence of a homogeneous gravitational field, will share all of these properties. The equations governing spacetime remain the same with the replacement of the term gx by the gravitational potential Φ. This is the most important part of the argument for it contains the crucial stepping from inertial fields to gravitational fields.

Step 3 We generalise the laws governing the behaviour of homogeneous gravitational fields to recover those governing arbitrary static gravitational fields. This step seems natural and unproblematic, for it only involves the relaxation of the condition that the gravitational potential Φ be given by Φ = gx and allowing it to be an arbitrary function of space coordinates, x,y and z. Thus spacetime, in the presence of a static gravitational field, is governed by the metric

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 = \left(c_L + \frac{\Phi}{c_L}\right)^2 dt^2 - dx^2 - dy^2 - dz^2 \qquad (5.2c)$$

and the isotropic speed of light, measured with coordinate times and distances, is given by

$$c = c_L + \frac{\Phi}{c_L} \qquad (5.3b)$$

For this reason, it is possible to use this speed in the place of the gravitational potential, an option that Einstein takes up. It also follows from the metric (5.2c) that the time read by resting clocks differs from coordinate time, whilst three dimensional space remains Euclidean.

This presentation of Einstein's argument has been greatly simplified by the use of non-Lorentzian metrics to characterise the properties of spacetime. Einstein, of course, did not introduce such metrics until after he had completed work on his scalar theories. We can readily confirm that the arguments used by Einstein in this 1907-1912 period actually corresponded with the one outlined above by examining his relevant papers:

The 1907 Jahrbuch Article [19a]

In the 1907 Jahrbuch...article, Einstein began his analysis of the kinematics of acceleration by arguing that any influence by acceleration on the length of rods must depend at least on the second power of that acceleration. Here he refers to the fact that if such effects were to depend on terms linear in acceleration, then an acceleration in one direction, which produced a contraction, would produce a dilation if turned in the other direction. This could only result from an unacceptable asymmetry in the background space. From this Einstein concluded that, as long as small accelerations are considered, acceleration will produce no effect on the length of rods. This corresponds to the recovery of the Euclidean nature of the spatial part of the metric (5.2b).

Einstein continued from this result to examine the behaviour of clocks, resting in an accelerated coordinate system. By comparing the behaviour of clocks in a reference inertial system with those in an inertial system instantaneously at rest with respect to the accelerated frame, he found the result (numbered "30" by him) for small τ and ξ

$$\sigma = \tau\left[1 + \frac{\gamma\xi}{c^2}\right]$$

19a. Einstein, "Über das Relativitätsprinzip...".

where his time σ corresponds to T here, τ to t, γ to g, c to c_L and ξ to x.

This result corresponds exactly to equation (5.4b) above. With it Einstein has recovered the features sufficient to characterise the inertial field described by the metric (5.2b). At this point Einstein did not go on to conclude that the speed of light remained isotropic but varied with position in the field according to equation (5.3a), although the results he had obtained would have been sufficient for him to do this. Rather, he arrived at an equivalent result at the conclusion of the paper by the indirect route of examining the effect of the field on Maxwell's equations.

Einstein did continue, however, with Step 2 of the above argument. He invoked the equivalence of uniform acceleration and homogeneous gravitational fields and concluded that all of the equations derived for the case of uniform acceleration applied also to homogeneous gravitational fields, provided one understood the quantity $\gamma\xi$ to stand for the gravitational potential Φ .

At this early stage of his work on gravitation, Einstein seemed loath to continue with Step 3 of the argument and generalise his results to the case of arbitrary static gravitational fields. He briefly entertained such an extension, however, in order to be able to apply his equation (30) to the inhomogeneous gravitational field of the sun and to conclude the existence of a gravitational redshift of the wavelength of light from the sun of about a two millionth part.

In his 1911 <u>Annalen der Physik</u> paper on gravitation, Einstein added little of relevance to this aspect of his work.[20] As we have seen, he tightened and extended certain aspects of his arguments and introduced the new result that a beam of starlight grazing the sun should be deflected by 0.83 seconds of arc.

The 1912 Theory

In 1912, in two consecutive papers, Einstein greatly extended his earlier gravitation theory without changing its foundations.[21] In his exposition of the foundations of this theory, the correspondence of his argument with that outlined above becomes especially clear. The relevant arguments of this 1912 theory have already been sketched out in Section 4.4 earlier.

It will be recalled that Einstein began by developing the kinematics of uniform acceleration in a gravitation free space. He considered the expansion of a shell of light as seen from an inertial system $\Sigma(\xi,\eta,\zeta,\tau)$, in which the speed of light is unity, and as seen from a uniformly accelerated system K(x,y,z,t), accelerating in the ξ direction of Σ. He postulated, presumably on the basis of the arguments of his earlier papers, that this process is governed in K by

$$dx^2+dy^2+dz^2-c^2dt^2=0$$

where c is the variable speed of light and, of course, in Σ by

$$d\xi^2+d\eta^2+d\zeta^2-d\tau^2=0$$

20. A. Einstein, "Über den Einfluss der Schwerkraft auf die Ausbreitung des Lichtes", Annalen der Physik,35(1911), pp.898-908.

21. A. Einstein, "Lichtgeschwindigkeit und Statik des Gravitationsfeldes", Annalen der Physik,38(1912), pp.355-69; "Zur Theorie des statischen Gravitationsfeldes", Annalen der Physik,38(1912), pp.443-58.

From this he concluded the invariance of the term

$$d\xi^2 - d\tau^2 = dx^2 - c^2 dt^2$$

These invariant terms clearly correspond to the metrics used to characterise the structure of spacetime in my summary of the argument.

By focussing on the invariance of these terms, Einstein then derived transformation equations which link the two coordinate systems for the case of small acceleration and small t and which correspond to equations (5.1b) and (5.3a) exactly:

$$\left.\begin{array}{l} \xi = x + \frac{ac}{2}t^2 \\ \eta = y \quad \zeta = z \quad \tau = ct \end{array}\right\} \qquad (4.3a)$$

where $c = c_0 + ax$ (4.3b)

and c_0 and a are constants. This last equation was numbered (5) by Einstein and asserted that the speed of light would vary linearly with coordinate distance in the direction of the acceleration.

With this Einstein had completed the recovery of the kinematics of uniform acceleration required by Step 1. From this he proceeded immediately to Steps 2 and 3 by constructing a field equation for an arbitrary mass free static field. This was accomplished in a single crucial paragraph:

> "From the earlier work it has already emerged, that in a static gravitational field a relationship exists between c and the gravitational potential, or in other words, that the field is determined through c. In those gravitational fields, which correspond to the acceleration fields considered in § 1, the equation
>
> $$\Delta c = \frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} + \frac{\partial^2 c}{\partial z^2} = 0$$
>
> is satisfied from (5) and the principle of equivalence and this suggests that we regard these equations as

valid in every mass free static gravitational field. In any case, this equation is the simplest compatible with (5)." 22

Here Einstein first recognised that homogeneous gravitational fields were also to be characterised by a speed of light which varies linearly with distance in the direction of the field (his equation (5)). This he noted was consistent with the field equation

$$\nabla^2 c = \frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} + \frac{\partial^2 c}{\partial z^2} = 0 \qquad (5.5a)$$

which he then assumed to govern arbitrary static and source mass free gravitational fields as well. This equation was later extended naturally to cover the case of a static field in the presence of source masses.

With this we can see that Einstein has arrived at a gravitation theory which is equivalent to the one arising from the assumption that spacetime is governed by the metric (5.2c), where the speed of light c must satisfy the appropriate field equation. In Appendix C, I confirm that essential features of the remainder of the theory which Einstein proceeded to develop, do indeed correspond to the behaviour of such a spacetime and the matter contained within it.

---

22. Aus der früheren Arbeit geht schon hervor, dass im statischen Gravitationsfeld eine Beziehung zwischen c und dem Gravitationspotential existiert, oder mit anderen Worten, dass das Feld durch c bestimmt ist. In demjenigen Gravitationsfelde, welches dem im § 1 betrachteten Beschleunigungsfelde entspricht, ist nach (5) an dem Äquivalenzprinzip die Gleichung

$$\Delta c = \frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} + \frac{\partial^2 c}{\partial z^2} = 0$$

erfüllt, und es leigt die Annahme nahe, dass wir diese Gleichung als in jedem massenfreien statischen Gravitationsfelde gültig anzusehen haben. Jedenfalls ist diese Gleichung die einfachste mit (5) vereinbare." Einstein, "Lichtgeschwindigkeit...",p.360.

5.4 The Flatness of Space

Einstein's postulation and use of the equivalence of acceleration and gravitation in the 1907 - 1912 period had been remarkably successful. In the theories which he drew from it in this period, he was able to recover a number of results and effects which would later feature prominently in his general theory of relativity. I will return to analyse this in greater detail before concluding the chapter. For the moment, it is more convenient to examine the limitations of Einstein's use of this equivalence, from the point of view of his final theory.

To begin with, we can foresee where Einstein's general strategy of inferring the structure of gravitational fields from that of inertial fields is likely to go astray. From the discussion of Section 5.1 , it is clear that inertial fields are capable of mimicking the gross features of gravitational fields only. Thus we would expect Einstein to have early success in recovering these gross features of gravitational fields, but to meet with difficulties in the recovery of the finer structure of gravitational fields.

To follow this through in more specific terms, consider Einstein's argument as outlined in Figure 5.1 . From the point of view of the general theory of relativity, the argument contains a disasterous flaw. As the crucial step of the argument, Einstein assumed that the postulation of the equivalence of uniform acceleration and homogeneous gravitational fields would act as a bridge which would enable him to pass from inertial to gravitational fields. However, in the general theory of relativity, such homogeneous gravitational fields would not be regarded as true gravitational fields at all. For, in their presence, spacetime would still remain flat. The acceleration of test bodies in the field could not be regarded as directed towards any specific source masses but to owe its origin to the acceleration of the coordinate system. The postulation of the equivalence of inertial fields produced by uniform acceleration and homogeneous gravitational fields would have to be regarded as no more than a definition, a decision to call such inertial fields by the name "homogeneous gravitational field".

However, it is exactly because this equivalence can be regarded as a definition that Step 2 of the argument must be considered an acceptable, if uninteresting, step, according to the general theory of relativity. Curiously, it is only with the apparently unproblematic Step 3 that the requirements of the general theory of relativity are violated. For this step does not comprise an innocuous generalisation of the behaviour of a special example of a static gravitational field to more general ones qualitatively similar to it. Rather, it involves a perilous leap from one type of field, "homogeneous gravitational fields", in whose presence spacetime remains flat and in which no specific masses can be pinpointed as sources, to a qualitatively different kind of field, general static gravitational fields, in whose presence spacetime is curved and in which bodies fall towards identifiable source masses.

This leap is a perilous one and with it came a conclusion which contradicts the general theory of relativity. From metric (5.2b), we can see that the spatial cross-section of the spacetime of a homogeneous gravitational field is flat. Einstein naturally carried this result over to arbitrary static gravitational fields. In the general theory of relativity, static gravitational fields are not spatially flat, even in the weak field case. In this case, they are described by the metric

$$ds^2 = (c_L^2 + 2\Phi)dt^2 - (1 - \frac{2\Phi}{c_L^2})(dx^2 + dy^2 + dz^2) \qquad (5.2d)$$

where Φ is small and a function of x,y and z only.[23] This of course compares to the metric for weak static fields which arises from Einstein's argument and is derived from the metric (5.2c) for the case of small Φ

$$ds^2 = (c_L^2 + 2\Phi)dt^2 - (dx^2 + dy^2 + dz^2) \qquad (5.2e)$$

The generalisation of Step 3 and, in particular, the associated derivation of the field equation in the 1912 theory, proceeds so naturally that is interesting to pinpoint exactly where the argument violates the requirements of the general theory of relativity. It will be recalled that Einstein had concluded that homogeneous gravitational fields are spatially flat with the speed of light varying linearly with distance in the direction of the field, x:

$$c = c_0 + ax \qquad (4.3b)$$

where c_0 and a are constants. This is a solution of the simple differential equation

$$\frac{\partial^2 c}{\partial x^2} = 0 \qquad (5.5b)$$

23. Fock, pp.179-84.

which readily suggest the isotropic generalisation

$$\nabla^2 c = 0 \tag{5.5a}$$

This equation is sufficiently weak to admit a wide range of solutions, which Einstein assumed corresponded to general source free static gravitational fields. In addition, Einstein implicitly carried over the conclusion that such fields would be spatially flat.

We arrive at a generalisation of equation (4.3b) which is in accord with the general theory of relativity if we recall that, in this theory, Einstein's "homogeneous gravitational fields" are not special instances of general static gravitational fields, but of fields in whose presence spacetime remains flat. This condition amounts to the vanishing of the components of the Riemann curvature tensor

$$R^{\alpha}{}_{\beta\gamma\delta} = 0$$

In Appendix C, I calculate the components of this tensor for a metric of the form of (5.2c). From this it follows that the above condition reduces to the set of six equations:

$$\frac{\partial^2 c}{\partial x^i \partial x^j} = 0 \tag{5.5c}$$

where i,j = 1,2,3.

This set of six equations, rather than equation (5.5a), is the isotropic generalisation of equations (4.3b) and (5.5b) which arises from the general theory of relativity. This set of equations is far more restrictive than equation (5.5a). Their most general solution only allows a linear dependence of c on the spatial coordinates:

$$c = c_0 + Ax + By + Dz$$

where A,B,D and c_0 are constants. If we retain the requirement that space remain Euclidean and note that this general solution for c can be reduced to equation (4.3b) by simple rotations of the spatial coordinate axes, we can see that the fields described by equations (5.5c) are not general source free static gravitational fields. Rather, they are inertial fields produced by uniform acceleration against a flat spacetime in an arbitrary direction or, in other words, homogeneous gravitational fields of arbitrary direction.

This conclusion is most satisfactory, for such fields are the only source mass free fields which the final theory will allow and which are spatially flat and have a metric of the form of (5.2c). We can confirm this directly by solving the source mass free field equations of the general theory of relativity

$$R_{\mu\nu} = 0$$

where $R_{\mu\nu}$ is the Ricci tensor for the case of such a metric. I calculate its components in Appendix C. From their values, it follows that this condition reduces to equation (5.5c), which, as we have just seen, solves to yield the required result.

5.5 Gravitational Red Shift and Light Deflection

Prior to 1913, in his work on gravitation, Einstein used the postulated equivalence of uniform acceleration and homogeneous gravitational fields as a bridge, which would enable him to pass from the accessible structure of inertial fields to gravitational fields and construct a new gravitation theory. As we have seen, according to the general theory of relativity, this "equivalence" cannot be regarded as such a bridge, for "homogeneous gravitational fields" are not seen to be true gravitational fields at all. In spite of this Einstein met with early success in this work, for inertial fields are capable of mimicking the gross features of gravitational fields, but not their finer structure.

So Einstein's scalar theory predicted a structure for weak static gravitational fields which is equivalent to the metric

$$ds^2 = (c_L^2 + 2\Phi)dt^2 - dx^2 - dy^2 - dz^2 \quad (5.2e)$$

whereas Einstein's final general theory of relativity leads to

$$ds^2 = (c_L^2 + 2\Phi)\,dt^2 - \left(1 - \frac{2\Phi}{c_L^2}\right)(dx^2 + dy^2 + dz^2) \quad (5.2d)$$

The gross properties of these fields are governed by the dt^2 coefficients, which are identical. So, for example, both predict essentially the same motions for bodies falling with small velocities. They disagree however, on the details of the finer structure of such weak static fields. Einstein's starting point in his argument had been an inertial field which was spatially Euclidean. He carried this assumption of spatial flatness over to static gravitational fields, on the basis of the postulated equivalence of acceleration and gravitation. This can be seen in the -1 coefficient of the $dx^2+dy^2+dz^2$ term of equation (5.2e). However, in the general theory of relativity, this equivalence does not extend to this detail of the structure of gravitational fields. Even weak static fields are not spatially Euclidean, according to this theory, as can be seen from an examination of the corresponding term of equation (5.2d).

The significance of this difference is well illustrated by the differing predictions concerning gravitational redshifting and deflection of light. Consider the weak static gravitational field produced by the sun's mass, M. From Newtonian correspondence, we have

$$\Phi = -\frac{GM}{r}$$

where r is the distance from the sun and G the gravitational constant.

For Einstein's scalar theory, this yields

$$ds^2 = \left(c_L^2 - \frac{2GM}{r}\right) dt^2 - dx^2 - dy^2 - dz^2 \qquad (5.2f)$$

and for the general theory of relativity

$$ds^2 = \left(c_L^2 - \frac{2GM}{r}\right) dt^2 - \left(1 + \frac{2GM}{c_L^2 r}\right)\left(dx^2 + dy^2 + dz^2\right). \qquad (5.2g)$$

This latter metric corresponds to the Schwarzschild metric in isotropic coordinates for the weak field case.[24]

From both metrics (5.2f) and (5.2g), it follows that the period dT of a light emitting atom on the surface of the sun at r = R will be related to coordinate time by:

$$ds = c_L dT = c_L \left(1 - \frac{GM}{c_L^2 R}\right) dt$$

From this it follows that the frequency of the light ν emitted will be shifted towards the red by the amount

$$\frac{\Delta\nu}{\nu} = -\frac{GM}{c_L^2 R}$$

which Einstein calculated to be $2\text{x}10^{-6}$ in both 1907 and 1911.[25] The agreement on this prediction between his early theory and later general theory of relativity stems directly from the fact that the calculation depends only on the time coefficient of the relevant metric.

---

24. R. Adler, M. Bazin and M. Schiffer, Introduction to General Relativity (Tokyo: McGraw Hill Kogakusha, 1975), p.199.

25. Einstein, "Über das Relativitätsprinzip...", p.459; "Über den Einfluss...". A fallacy in the traditional derivation of this result has been discussed in J. Earman and C. Glymour, "The Gravitational Red Shift as a Test of General Relativity: History and Analysis", Studies in the History and Philosphy of Science, 11 (1980), pp.175-214.

However, the two metrics yield different predictions for the angle of deflection of a beam of starlight which grazes the surface of the sun. I show in Appendix D that the angle of deflection derived from the metric (5.2g) of the general theory of relativity is given by

$$\Delta\theta = \frac{4GM}{c_L^2 R}$$

and that the time and space coefficients of the metric each contribute an equal half deflection of $\frac{2GM}{c_L^2 R}$ to this final result.

In the case of the metric (5.2f) of Einstein's scalar theory, which is spatially Euclidean, only the time coefficient contributes to the total deflection figure. Thus the deflection predicted will be only the half deflection of $\frac{2GM}{c_L^2 R}$. Einstein recovered this half deflection angle from his scalar theory in 1911 and calculated it to be 0.83 seconds of arc.[26]

By 1916, after the discovery of his 1915 field equations, Einstein recognised that his general theory of relativity did not allow space to remain Euclidean, even in this weak field case. The contribution of this extra effect led him to the full deflection of $\frac{4GM}{c_L^2 R}$, which he calculated to be 1.7 seconds of arc.[27]

We can see here quite clearly that the predictions made about gravitational light deflection are different in the case of the two theories because this prediction is far more sensitive to the finer structure of the gravitational field than are predictions about gravitational red shifts. In particular, the angle of deflection of

---

26. Einstein, "Über den Einfluss...".

27. A. Einstein, "The Foundation of the General Theory of Relativity" in The Principle of Relativity (New York: Dover, 1952), pp.160-4.

light by a gravitational field is strongly dependent on the curvature of the spatial cross-section of spacetime, whereas the degree of gravitational red shifting of light is unaffected by it.

The assumption that space remains Euclidean in the presence of a static gravitational field led to further difficulties for Einstein. It appears that he simply transplanted this conclusion from his 1912 scalar theory to his 1913 "Entwurf..." paper, in which he presented the framework of his general theory of relativity. For early in the "Entwurf..." paper, immediately after a brief survey of his 1912 scalar theory, he discussed the need for a new theory to generalise the existing theory of relativity and which would contain his 1912 theory as a special case. [28] On the next page he presented the Lorentzian metric of the spacetime of his special theory of relativity, where the "time-time" component of the metric, g_{44}, is given by the square of the speed of light. He continued:

> "The same type of degeneration is shown by static gravitational fields of the type previously considered, except in this case $g_{44}=c^2$ is a function of x_1,x_2,x_3." [29]

Stachel has argued that it was this misconception rather than a mathematical mistake which led to Einstein and Grossmann's notorious rejection, in the "Entwurf..." paper, of the Ricci tensor $R_{\mu\nu}$ as the basis of the gravitational field equations. [30] Calculation of

---

28. A. Einstein and M. Grossmann, "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation", Zeitschrift für Mathematik und Physik, 62 (1913), p.228.

29. Dieselbe Art der Degeneration zeigt sich bei dem statischen Schwerefelde der vorhin betrachteten Art, nur dass bei diesem $g_{44}=c^2$ eine Funktion von x_1,x_2,x_3 ist." Ibid., p.229.

30. J. Stachel, "Einstein's Search for General Covariance, 1912-1915", talk given at the Ninth International Conference on General Relativity and Gravitation, Jena, DDR, July 17, 1980. I am most grateful to Professor Stachel for providing me with a draft manuscript of his talk.

the components of the Ricci tensor for a weak field metric of the form of (5.2e) yields the following non-zero terms (See Appendix C):

$$R_{00} = \nabla^2\varphi \qquad R_{ij} = -\frac{1}{c^2}\frac{\partial^2\varphi}{\partial x^i \partial x^j} \qquad i,j = 1,2,3$$

If Einstein and Grossmann made such a calculation, they could have immediately eliminated the Ricci tensor as the basis of the field. For even the simple field equation

$$R_{\mu\nu} = 0$$

would only allow the gravitational potential φ to vary linearly with the distance coordinates. I will return to this question in a later chapter.

5.6 The Contribution of the Principle of Equivalence

In this final section, I seek to draw together the material of this chapter and the preceding one and to summarise the contribution of Einstein's postulation of the equivalence of acceleration and gravitation and the principle of equivalence to his discovery of the general theory of relativity. First I deal with the 1907 - 1912 period, prior to the emergence of the "Entwurf..." theory.

5.6.1 Prelude to the General Theory of Relativity

In 1907, Einstein felt that his special theory of relativity was incomplete, for it did not enable an extension of the relativity of motion to accelerated motion as well. He later recalled this and noted that even his awareness of Mach's work in the area"...provided no workable basis for a new theory". [31] However, at the same time,

---

31. A. Einstein, "Notes on the Origin of the General Theory of Relativity" in Ideas and Opinions (London: Souvenir, 1973), p.286.

out of his work on gravitation in relativity theory came a most provocative idea, the equivalence of acceleration and gravitation, an idea in which Einstein saw such a workable basis. He began work in a programme whose focus was this equivalence and whose ultimate goal was the generalisation of his earlier special theory of relativity. By exploiting this equivalence, he hoped to find a way of eliminating the absolute status which acceleration had enjoyed in earlier theories.

We have seen that, in terms of the final general theory of relativity, this programme contained a fundamental flaw. There is no thoroughgoing equivalence of acceleration and gravitation of the type which Einstein invoked. However, since the gross effects of gravitation and acceleration are the same, Einstein was able to recover a number of important results which were to figure prominently in his final general theory of relativity, without the need for the complex theoretical and conceptual apparatus of this final theory.

First Einstein could conclude that special relativity could only hold as a boundary case. For in the presence of a gravitational field or in an accelerated frame of reference the light postulate would no longer hold, the speed of light now being variable. Contained within this was a second equally important conclusion: the gravitational potential was now to be represented by the variable speed of light. This amounted to a shift in the way that gravitational fields were dealt with formally. They were no longer seen as entities contained within the uniform containers of space and time. Rather they were seen to be irregularities in the uniformity of the container itself. This, of course, foreshadows the treatment of gravitational fields as the curvature of spacetime in the final general theory of relativity. As we have seen, Einstein was even able to recover predictions closely associated with this concerning the gravitational deflection and red shifting of light. The amount of red shift corresponded exactly to that predicted by the final theory, whereas the angle of light deflection was exactly one half of the final theory's prediction.

Finally the principle of equivalence played an important role in bringing Einstein to acknowledge a most significant result. That is, space and time coordinates are no longer to be equated directly with the measurements of rods and the readings of clocks or, in other words, they are to lose their direct metrical significance. This is a result that is fundamental to the final general theory of relativity. For it is only through this that it is possible to introduce arbitrary coordinate systems. In the theory space and time coordinates can only be related to the measurements of rods and the readings of clocks by the mediation of the metric tensor. This question has not been discussed here explicitly so far. Therefore I will now discuss it at slightly greater length than the other issues of this section.

Einstein found this most crucial of results difficult to accept, as he was to recall later in his Autobiographical Notes.[32] To understand this we need only note that the key step of his 1905 paper on special relativity was the insistence on an antithetical requirement, that space and time coordinates, and the time coordinate in particular, are to be given directly by the results of measuring operations. The conclusion that space and time coordinates must

32. A. Einstein, Autobiographical Notes (La Salle and Chicago, Illinois: Open Court, 1979), p.63.

forfeit this direct metrical significance, if a generalisation of the special theory of relativity is to be constructed, seems to have been forced on Einstein only by the accumulated weight of a number of considerations.

In a recent paper, Stachel has argued that Einstein's analysis of the problem of the rotating disk was instrumental in leading Einstein to this conclusion. [33] For well before the 1913 "Entwurf..." theory, Einstein had recognised that the geometry yielded by measuring rods in the stationary field of a rotating frame of reference could not be Euclidean.

In addition to Stachel's point, we should note that Einstein's attempts to come to terms with the structure of the gravitational fields made accessible by his postulated equivalence of gravitation and acceleration also led him to this same conclusion that space and time coordinates must forfeit their direct metrical significance. In a 1912 discussion of his then current gravitation theory and the possibility of developing an extended theory of relativity from it, he wrote:

> "Admittedly the way to this goal seems certainly to be more difficult. One can see already from the highly specialised case of the gravitation of resting masses treated hitherto, that space-time coordinates will forfeit their simple physical meaning, and it is still not foreseen what form the general spacetime transformation equations could have." [34]

The forfeiture of the simple physical meaning of spacetime coordinates, described by Einstein as arising in his theory of static gravitational fields, refers to the behaviour of the time coordinate in that theory.

---

33. J. Stachel, "Einstein and the Rigidly Rotating Disk" in General Relativity and Gravitation: A Hundred Years after the Birth of Einstein, Vol.1, ed. A. Held (New York: Plenum,n.d.-1979?), pp.1-15.

34. A. Einstein, "Relativität und Gravitation", Annalen der Physik, 38 (1912), pp.1063-4. For German text, see Section 4.4 .

As early as 1907, Einstein had found that there were "times" in a gravitational field which have no direct metrical significance. On the one hand there was local time, which was the time read by clocks. This had a purely local significance. To deal with global quantities and effects, he found that he had to introduce another time, in effect a non-metrical coordinate time, which was not measured directly by clocks, but which could be constructed from their readings.[35]

Later, in his 1913 "Entwurf..." paper, Einstein presented an outline of his new general theory of relativity and noted that in it space and time coordinates had lost their direct metrical significance. However, in noting this in his first exposition of the theory, he did not turn for support to the rotating disk thought experiment, as he was to at other times.[36] Rather, he turned again to his earlier gravitation theory and its treatment of time and the time coordinate:

> "From the above, one can already conclude that no such simple relationship can exist between the space-time-coordinates x_1, x_2, x_3, x_4 and the measurements obtained by means of measuring rods and clocks as does in the old relativity theory. This was shown with respect to time already with static gravitational fields..." [37]

A footnote appended to this last sentence directed attention to Einstein's 1911 Annalen der Physik paper. This passage again confirms the importance of Einstein's experience with the static gravitational fields of his earlier theory in his arriving at this conclusion.

---

35. Einstein, "Über das Relativitätsprinzip...".

36. Einstein, "The Foundation...", pp.115-7.

37. "Aus dem Früheren kann man schon entnehmen, dass zwischen den Raum-Zeit-Koordinaten x_1, x_2, x_3, x_4 und den mittelst Massstäben und Uhren zu erhaltenden Messergebnissen keine so einfachen Beziehungen bestehen können, wie in der alten Relativitätstheorie. Es ergab sich dies bezüglich der Zeit schon beim statischen Schwerefelde..." Einstein, "Entwurf...", p.230.

5.6.2 The General Theory of Relativity

Throughout the development of the general theory of relativity, a fundamental postulate of the theory was always taken to be one form or another of a principle of equivalence. In 1913, with the emergence of the first outline of the theory in the "Entwurf..." paper, we have seen that Einstein took the principle to be the familiar assertion of the indistinguishability of the effects of uniform acceleration and homogeneous gravitational fields. We have seen that, in Einstein's hands, this form of the principle did not seem to make any specific contribution to the content of the theory and that, by 1918, Einstein had come explicitly to a quite different statement of the principle and its significance in the theory.

In the light of the analyses of this chapter, this shift is most appropriate. I have argued that, in the general theory of relativity, the equivalence of uniform acceleration and homogeneous gravitational fields can amount to no more than a definition, a statement that certain inertial fields are to be labelled "gravitational".

In 1918 and 1920, Einstein addressed himself to the question of whether inertial fields in general could be regarded to be real gravitational fields.[38] He presented a number of related considerations. To begin with, he questioned the utility of the

---

38. A. Einstein, "Dialog über Einwände gegen die Relativitätstheorie", Die Naturwissenschaften, 6 (1918), pp.697-702; "Meine Antwort auf die antirelativitätstheoretische G.m.b.H.", Berliner Tageblatt und Handelszeitung, 27th August 1920, pp.1-2.

distinction real-unreal. He did not see that one should regard inertial fields as ficticious fields simply because they could be transformed away by a shift to a new coordinate system. For example, the reality of kinetic energy was not doubted, even though it shared this same property. Further, he noted that it was not an a priori necessity that every gravitational field should be seen to be generated by masses, but only a concept of Newtonian theory which need not be carried over to the general theory. In any case, however, he noted his belief that inertial fields could be seen indirectly to have their sources in all the masses of the universe, although this was not universally agreed to by all the supporters of relativity theory. This belief was associated with what he termed "Mach's principle" and will be dealt with in the following chapter.

In effect, all that Einstein was doing here was elucidating his definition of a gravitational field. So, for him, a defining characteristic of a gravitational field was not that one should be able to identify its specific source masses. Further, he had argued the fact, that some fields could be transformed away amounted to the vanishing of the "components of the gravitational field" or, in other words, the components of the affine connection, with a suitable choice of coordinate system. This contains an implicit definition. Specifically, he defined there to be no gravitational field at a point in spacetime in a given coordinate system if these components were all zero. An alternative definition used in this chapter is that there exists a "true" gravitational field at a point in spacetime if and only if all the components of the Riemann curvature tensor are non-zero.

Certainly each definition has its own conceptual advantages, but neither can be right or wrong. They are just definitions. However, the choice of definition here does not affect whether the postulated equivalence of uniform acceleration and homogeneous gravitational fields can make a contribution to the content of the final theory. The answer to this question remains the same. This equivalence does not provide a crucial bridge between fields which are qualitatively different. The equivalence deals only with a limited type of field, those associated with flat spacetimes, and cannot be guaranteed to provide any access to the vast majority of fields of great interest, those associated with curved spacetimes.

Whilst this equivalence cannot be said to have made a direct contribution to the content of the final theory, it did make a contribution to the development of the theory. We have seen in the last chapter how the equivalence acted as a most useful heuristic. It suggested to Einstein that an extension of the principle of relativity was within his grasp and provided what he saw as a simple example of how this could be done in a special case. In particular, it led Einstein to see that the inclusion of gravitation into his theory was the key to the realisation of this extension.

Finally this equivalence led Einstein to a form of the principle of equivalence which is of paramount importance to the foundations of the general theory of relativity. The equivalence had originally arisen through an attempt to capture the fundamental significance of the distinctive property of gravitational fields, that the motion of a freely falling body is to a very large extent independent of the nature and state of the body. From this independence, it follows that the trajectories of freely falling bodies form a unique set of lines, which fill spacetime. This special property of gravitation finds expression in the general theory of relativity in the identification of this set of lines with the set of geodesics of a metrical spacetime.

In the last chapter we saw how Einstein came to recognise that the significance of the principle of equivalence lay in this result. He reformulated the principle as the statement that inertia and weight are of the same nature. This found expression in the fact that the theory treated both inertial and gravitational phenomena through the structure of spacetime. Specifically both are represented by the metric tensor.

5.7 Conclusion

In 1907 Einstein focussed his attention on the independence of the motion of a freely falling body from its nature and state. In an attempt to capture the significance of this distinctive property of gravitation, he postulated the equivalence of uniform acceleration and homogeneous gravitational fields. He hoped that he would be able to construct an extension of his special theory of relativity on the basis of an exploitation of this equivalence. He began work on a new relativistic theory of gravitation, in which he used this equivalence as the crucial bridge which would enable him to pass from the accessible structure of inertial fields to the more obscure structure of gravitational fields.

Within this theory he was able to recover important results and effects of the final general theory of relativity, without the use of that theory's elaborate theoretical and conceptual machinery. He concluded that special relativity could not hold in the presence of a gravitational field, for in the presence of such a field, the speed of light would no longer be constant and could even stand for the gravitational potential itself. This amounted to the formal incorporation of gravitational fields into the structure of space and time, a result which is crucial to the final general theory of relativity. He was also able to conclude that gravitational fields would deflect light rays and slow clocks and, in conjunction with other considerations, came to accept that space and time coordinates must lose their direct metrical significance, if an extension of the special theory of relativity is to be realised.

However, according to his final general theory of relativity, there is no thoroughgoing equivalence of acceleration and gravitation of the kind which Einstein invoked. Indeed the so-called "homogeneous gravitational fields", whose effects were postulated to be equivalent to those of uniform acceleration, would no longer be regarded to be gravitational fields at all. But, within the theory, inertial fields do mimic certain gross features of gravitational fields and this accounts for Einstein's early success with his scalar theory of static gravitational fields. It even enabled him to recover the final theory's exact predicted value for the degree of red shifting of light from the sun in his earlier theory.

Inertial fields differ from gravitational fields when it comes to their finer structure. Thus Einstein came to the conclusion that static gravitational fields are spatially Euclidean, in contradiction to his final theory. This was directly responsible for his recovery of only half the final theory's predicted angle of deflection of a ray of starlight grazing the sun. Einstein also carried this conclusion over to the earlier versions of his general theory of relativity. This may have played a role in the premature 1913 rejection of the final 1915 generally covariant field equations.

In spite of this, a principle of equivalence was still to play a significant role in the development of the general theory of relativity from the time of the "Entwurf..." paper onwards. In the form of an assertion of the equivalence of uniform acceleration and homogeneous gravitational fields, the principle provided a suggestive example of how an extension of the range of frames of reference covered by the principle of relativity may be realised and showed that such an extension would require the incorporation of gravitation into the new theory.

Einstein came to realise that the function of justifying an extension of this range of frames of reference was completely taken over by the explicit invocation of the generalised principle of relativity. The only contribution left to be made by the principle of equivalence lay in the latter result, that gravitation was to be incorporated into the new theory and, as a result, into the structure of spacetime itself. Thus Einstein reformulated the principle as the statement that inertia and weight are of the same nature and noted that this found expression in the theory in the representation of both inertial and gravitational phenomena by the same theoretical machinery associated with the metric tensor.

CHAPTER 6

THE GENERAL RELATIVITY OF MOTION

6. THE GENERAL RELATIVITY OF MOTION

From his earliest speculations on gravitation in relativity theory, Einstein made it clear that he hoped they would yield an extension of the principle of relativity. The immediate problem to be solved in seeking such an extension, in so far as it asserted the physical equivalence of accelerated as well as inertial frames of reference, was the problem of inertial forces. The existence of these forces in accelerated frames of reference distinguished them from those in uniform motion and seemed to provide an absolute criterion for distinguishing acceleration from uniform motion. Einstein's 1907 insight was the recognition that these inertial forces behave in a very similar way to gravitational forces and he hoped, therefore, that an extension of the principle of relativity was to be found in a gravitation theory based on the postulation of the equivalence of gravitation and acceleration.

By 1912, Einstein had developed a sophisticated gravitation theory on the basis of his original insight. However, he seemed to be no closer to an extension of the principle of relativity than he had been in 1907. The focus of the extension was still his original idea that uniform acceleration and homogeneous gravitational fields were indistinguishable, but it had not advanced beyond this. If anything it had been weakened by Einstein's 1912 conclusion that the equivalence could only hold for infinitesimal regions. Whilst this equivalence was very suggestive, the question of exactly how the principle of relativity was to be extended remained unclear.

In Chapter 4, we saw how, in the same year, Einstein was finally able to recover the promise of a more complete account of the origin of inertial forces in his 1912 gravitation theory. Focussing on the idea that inertial forces were really gravitational in nature, he recovered two effects: the inertia of a body is increased by the piling up of other masses in its neighbourhood and an accelerated shell of matter will drag a centrally placed test body along with it. This suggested to Einstein that inertial forces in their entirety might be due to a gravitational interaction with all the other masses of the universe which paralleled the electromagnetic inductive interactions of accelerated charges. With this idea, which he attributed to Mach, Einstein was now able to present a picture of how the general relativity of motion could be established. Inertial forces did not arise through the intereaction of bodies in certain states of motion in space. This would violate the demand of the general relativity of motion. Rather they arose through a gravitational inter-action between all the bodies of the universe alone.

The 1912 paper which describes this effect is an important link between Einstein's work on gravitation before and after 1913. For this account of the origin of inertia and the general relativity of motion was to lie at the heart of Einstein's understanding of the conceptual foundations of his general theory of relativity, especially in its earlier years. Indeed Einstein was to become so convinced of the correctness of this account that he came to stress its origins as lying not in an empirical shortcoming of earlier theories of mechanics but in a fundamental epistemological defect contained within them.

I now turn to the task of analysing Einstein's understanding of the foundations of this theory after 1913 and their relationship to Mach's critique of Newtonian mechanics.

6.1 The Relativity of Inertia

By 1916, it had become Einstein's practice to introduce expositions and discussions of his general theory of relativity with careful arguments for the general relativity of motion and the related requirement that inertial forces arise from an interaction between the bodies of the universe. [1] This is the appraoch that today is still often associated with the name of Einstein. However, when the basic structure of the theory emerged in 1913, Einstein was by no means as confident in his assertions of the relationship between his new theory and the general relativity of motion. In his 1913 "Entwurf..." paper, he did not argue for the necessity of an extension of the principle of relativity on the basis of the epistemological arguments which were to figure prominently later. Rather he observed that his 1912 gravitation theory, by introducing a variable speed of light, had broken through the bounds of his earlier special theory of relativity. This led him to conclude:

> "If therefore - which cannot be doubted - the relativity principle is to be retained, then we must generalise relativity theory in such a way, that it contains the previously suggested theory of static gravitational fields in its elements as a special case." [2]

---

1. See, for example, A. Einstein, "The Foundation of the General Theory of Relativity", The Principle of Relativity (New York: Dover, 1952), pp.109-64.

2. "Soll also - woran nicht zu zweifeln ist - das Relativitätsprinzip aufrecht erhalten werden, so müssen wir die Relativitätstheorie derart verallgemeinen, dass sie die im vorigen in ihren Elementen angedeutete Theorie des statischen Schwerefeldes als Spezialfall enthält." A. Einstein and M. Grossmann, "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation", Zeitschrift für Mathematik und Physik, 62 (1913), p.228.

Here Einstein did not make clear the extent to which the relativity principle is to be generalised. We find in the theory which he went on to describe that the laws governing the action of gravitation on material processes take on a generally covariant form, suggesting a complete generalisation of the principle of relativity to all states of motion. [3] However, the field equations of the theory are not generally covariant. He did not rule out the possibility that such generally covariant field equations might exist and speculated that they may involve higher orders of the derivatives of the metric tensor than the second. But he believed that with the then current state of knowledge of gravitation, further discussion of the question would only lead them astray. [4]

It appears natural to assume that Einstein's hesitancy in discussing the general relativity of motion in his theory stems from this incompleteness and that this led to his uncertainty on the extent to which the general relativity of motion was contained within the theory. So, for example, he finished his part of the "Entwurf..." paper with an argument against the acceptability of a scalar gravitation theory. He concluded with the following remarks:

> "Certainly I must admit that for me the most effective argument for the case that such a theory should be rejected rests on the conviction that relativity should not only hold for linear orthogonal substitutions but for a much wider group of substitutions. But we are not yet entitled to make good this argument since

---

3. Ibid., P.229.

4. Ibid., pp.233-4.

> we are not in a position to find the (most general) group of substitutions which belong to our gravitation equations."[5]

In his Vienna address later in the year, Einstein took a similar approach to the question of the general relativity of motion in his theory and the failure of his field equations to be generally covariant.[6] In particular he recognised that a completely satisfactory gravitation theory would need generally covariant field equations.[7] In the ensuing discussion of his address, Mie challenged Einstein on the principle of general relativity and suggested that it had no physical meaning. In his retrospective reply, Einstein admitted that he did not regard his theory as consistent with a completely general principle of relativity:

> "In my theory the principle of relativity in this most general sense is likewise not realised. The conservation laws lead to a far reaching specialisation of the reference system, as I explained in the lecture..."[8]

---

5. "Ich muss freilich zugeben, dass für mich das wirksamste Argument dafür, dass eine derartige Theorie zu verwerfen sei, auf der Überzeugung beruht, dass die Relativität nicht nur orthogonalen linearen Substitutionen gegenüber besteht, sondern einer viel weiteren Substitutionsgruppe gegenüber. Aber wir sind schon deshalb nicht berechtigt, dieses Argument geltend zu machen, weil wir nicht imstande waren, die (allgemeinste) Substitutionsgruppe ausfindig zu machen, welche zu unseren Gravitationsgleichungen gehört." Ibid., p.244.

6. A. Einstein, "Zum gegenwärtigen Stande des Gravitationsproblems", Physikalische Zeitschrift, 14 (1913), pp.1249-66.

7. Ibid., p.1257.

8. "Nach meiner Theorie ist das Relativitätsprinzip in diesem allgemeinsten Sinne ebenfalls nicht erfüllt. Die Erhaltungssätze fuhren zu einer weitgehenden Spezialisierung des Bezugssystems, wie ich im Vortrage ausgeführt habe..." Ibid., p.1264.

The conservation laws referred to would now be written as

$$\left(\sqrt{-g}\,T_{\sigma}^{\nu}+\sqrt{-g}\,t_{\sigma}^{\nu}\right)_{,\nu}=0$$

where g is the determinant of the metric tensor, T_{σ}^{ν} the stress-energy tensor of non-gravitational matter and t_{σ}^{ν} the stress-energy tensor of the gravitational field. Einstein recognised that this latter quantity was not a generally covariant tensor and therefore concluded that the above conservation law, and with it his field equations, could not be generally covariant, but must involve a specialisation of the coordinate system.[9]

However, thus far in 1913, Einstein did not take this as proof that generally covariant field equations were physically unacceptable. In the writing of the body of these two papers, it is clear that Einstein had not yet constructed the argument which was to prove to his satisfaction that generally covariant field equations would violate causality. This notorious argument was added as an appendix to his "Entwurf..." paper and alluded to in a footnote in the Vienna address which was added after the bulk of the text had been completed.[10] We shall see later that Einstein's more detailed discussions of the general relativity of motion in the foundations of his theory appeared after this argument had convinced him that generally covariant field equations would be inappropriate for his theory - although Einstein did not mention any connection here.

---

9. Ibid., pp.1257-8.

10. Einstein, "Entwurf...", pp.260-1; Einstein, "Zum gegenwärtigen...", p.1257. For discussion of Einstein's struggle with the covariance properties of the field equations see J. Earman and C. Glymour, "Lost in the Tensors: Einstein's Struggles with Covariance Principles 1912-1916", Studies in the History and Philosophy of Science, 9 (1978), pp.251-78.

In spite of his early hesitancy, there can be no doubt that throughout this period Einstein was committed to achieving the general relativity of motion within his theory. This can be seen clearly in his treatment of the question of the origin of inertia. We have already seen how Einstein's 1912 speculations on the origin of inertia relate to the general relativity of motion. Einstein repeated the results of his 1912 analysis of the question in the introduction to his "Entwurf..." paper and then, in his Vienna address, showed how similar effects could be recovered from his new theory.

In that paper, he named the hypothesis that inertial forces arose through an interaction between an accelerated point mass and all other point masses the "hypothesis of the relativity of inertia".[11] He devoted a section to the hypothesis and there derived three effects consistent with it from his new theory.[12] First, on the basis of the lowest order weak field approximation, Einstein derived the equations of motion of a point mass. From these he showed that the inertial mass of a given body depends on the gravitational potential in such a way that the accumulation of masses in the region of the test body increases its inertia. This result turns out to be formally identical to the corresponding result in Einstein's 1912 theory. For Einstein's 1913 field equations yielded a weak field metric corresponding to (5.2e), rather than the metric (5.2d) of the final 1915 field equations. We have seen in Chapter 5 that this metric leads to equations of motion identical with those of the 1912 scalar theory.

---

11. Einstein, "Zum gegenwärtigen...", pp.1260-1.

12. Ibid., pp.1260-2.

To recover the other two effects, Einstein considered the weak field which arose from considering quantities of the next order of smallness. From this he recovered a set of equations formally identical to those of (a modern version of) Maxwell's electrodynamics, up to signs and multiplicative constants. The non-zero components of the metric tensor were g_{14}, g_{24}, g_{34} and g_{44}. The "time-time" component, g_{44}, corresponded to the electromagnetic scalar potential and the three remaining components to the vector potential.

These equations predicted, as a counterpart to the effect of electromagnetic induction, that an accelerating group of charges will induce forces on an unaccelerated test body which will drag it along in the direction of the acceleration. This, of course, brings to mind the case of a test mass accelerating with respect to the masses of the universe and suggests the possibility that the inertial forces it experiences may well arise totally from such an inductive effect.

Further, Einstein argued that the term corresponding to magnetic force in the equation of motion led to the existence of a Coriolis field inside a rotating spherical shell of matter and that this would cause precessional motion in the oscillation of a pendulum within the shell. This, of course, brings to mind the well known Foucault pendulum experiment in which the diurnal rotation of the earth is demonstrated by the appropriate precessional motion in the oscillation of a pendulum. Einstein's account here suggests an alternative explanation of the experiment. One might consider the earth to be nonrotating and the Coriolis forces which lead to the precession as induced by the diurnal rotation of all the extra-terrestrial matter of the universe.

These three effects are best known today through Einstein's treatment of them in his 1921 Princeton lectures, published as the book, The Meaning of Relativity.[13] Prior to this Thirring and Lense had given a detailed analysis of the third effect in 1918. In extending Einstein's analysis, in this case for the final 1915 field equations, they were able to recover terms corresponding to centrifugal forces, as well as Coriolis forces, in the fields of rotating sources.[14]

Einstein's 1921 treatment of these three effects differs from the original 1913 treatment. In particular, the use of his 1915 field equations leads to a different weak field approximation. But nevertheless, he is still able to recover a set of equations that closely parallel the equations of Maxwell's electrodynamics, from which the second and third effect directly follow. However, unlike his 1913 treatment, the first effect, that the inertia of a test body increases with the accumulation of other masses in its neighbourhood, was recovered directly from these equations as well, from a term which has no counterpart in the 1913 equations.

The anlysis of these effects has been a subject of some controversy. Brans, for example, has argued that the first effect stems only from Einstein's choice of coordinates.[15] Also, there has been

---

13. A. Einstein, The Meaning of Relativity (Chapman and Hall: London, 1976), pp.95-98.

14. H. Thirring, "Über die Wirkung rotierender ferner Massen in der Einsteinschen Gravitationstheorie", Physikalische Zeitschrift, 19 (1918), pp.34-39; 22 (1921), pp.29-30; J. Lense and H. Thirring, "Über den Einfluss der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinsche Gravitationstheorie," Physikalische Zeitschrift, 19 (1918), pp.157-63; H. Thirring "Über die formale Analogie zwischen den elektromagnetischen Grundgleichungen erster Näherung", Physikalische Zeitschrift, 19 (1918) pp.204-5.

15. C. Brans, "Absence of Inertial Induction in General Relativity", Physical Review Letters, 39 (1977), pp.856-7.

criticism of the consistency of Einstein's handling of small order quantities in the derivation of these effects. In any case, perhaps apart from the first effect, there is general agreement that weak field approximations do yield the type of forces consistent with Einstein's postulated relativity of inertia. [16]

The prediction of these effects by the theory is often taken to justify the belief that the theory entails the relativity of inertia. Whatever justification these effects may provide is unclear. The forces involved are very much smaller than known inertial forces. A more complete justification would require a problematic integration of the effects over all the masses of the universe. Rather than do this, Einstein concluded his account of the effects in his 1913 Vienna address with the comment that they were so weak that we could not hope to detect them in terrestrial experiments or in astronomy. Further, since the effects stem from a weak field approximation which yields equations similar to those of Maxwell's electrodynamics, there is no reason to believe that Einstein's theory is the only gravitation theory which would make such predictions. The feature of these equations involved in the prediction of inductive effects seems to be related not so much to the question of the general relativity of motion, but to the fact that they admit a maximum speed at which a disturbance might propagate within the field.

---

16. For discussion in this area and of the relevant literature, see N. J. Golden, "Some Aspects of Mach's Principle within the Theory of General Relativity", Diss. Univ. Wyoming 1971.

However, these considerations would not have been a great concern to Einstein in 1913. Then he was still not sure of the extent to which his theory contained the general relativity of motion and, as we have seen, the cause of this uncertainty was his field equations. It is clear that he believed that his theory had brought him closer than ever before to a realisation of this goal and, presumably, he hoped that his 1913 field equations closely approximated the generally covariant equations. Thus, if anything, it would have been encouraging that the three effects stemmed from a weak field approximation and not from some feature peculiar to his 1913 field equations. In particular, they stemmed from the weak field equation

$$\Box^2 g^*_{\mu\nu} = \kappa T_{\mu\nu}$$

where $g^*_{\mu\nu}$ represents the deviation of the components of the metric tensor from Lorentzian values, $T_{\mu\nu}$ the stress-energy tensor for non-gravitational matter and κ a constant. In 1913, Einstein assumed that the field equations should reduce to this form in the weak field case, since it is the simplest non-generally covariant second rank tensor generalisation of Poisson's equation.[17] (The 1915 generally covariant field equations did not reduce to this form, although, as we have seen, this did not impair the recovery of the appropriate effects from the theory. I will argue later that Einstein's misconception here about the form of the weak field equations played a significant role in the premature 1913 rejection of the 1915 generally covariant field equations.)

Thus the recovery of these three effects from his theory would have reassured Einstein that his theory had brought him close to

---

17. See, for example, Einstein, "Entwurf...", p.235.

his goal of the general relativity of motion. In a brief account of the theory written in 1913, he went so far as to describe the recovery of these effects from his theory as "...one of the most important supports of the outlined theory". [18]

Einstein was also able to use the recovery of these effects as a powerful selection criterion in 1913. In his Vienna address he outlined what he considered to be his theory's most significant competitor, Nordström's theory of gravitation. Clearly Einstein could not accept this theory since it was by no means a theory which realised the general relativity of motion. But, in 1913, Einstein was in no position to publicly reject it on these grounds for, as he was forced to admit, his own theory suffered from the same defect. However, he was able to dismiss Nordström's theory on grounds closely related to this question, for he found that he could not recover effects consistent with the relativity of inertia from it. He concluded his exposition of Nordström's theory with the comment:

> "In conclusion we can say that Nordström's scalar theory, which retains the postulate of the constancy of the speed of light, complies with all conditions, which can be placed on a theory of gravitation with the current state of experience. The only condition which remains unsatisfied, is that according to this theory the inertia of bodies is certainly influenced by the remaining bodies, but does still not appear to be caused by them, for according to this

---

18. "...eine der wichtigsten Stützen der skizzierten Theorie." A. Einstein, "Physikalische Grundlagen einer Gravitationstheorie", Naturforschende Gesellschaft Vierteljahrsschrift (Zürich), 58 (1913), p.290.

theory the inertia of a body becomes greater, the further we separate the remaining bodies from it"[19]

This latter result had been established by Einstein in the body of his exposition of Nordström's theory.

Thus, in summary, we can see that the relativity of inertia was of the greatest importance to Einstein in the earliest days of his new theory. For then he was still uncertain of the extent to which his new theory realised the general relativity of motion. The hypothesis and the weak field effects consistent with it provided confirmation of his belief that he had come very close to formulating a theory of general relativity and, in particular, provided him with a criterion for dismissing what he saw as his theory's only serious competitor, Nordström's theory.

6.2 The Epistemological Defect

By 1914, Einstein had more confidence in his theory. In March he wrote to his friend Besso and described his current work[20] on the field equations of the theory and his discovery of the condition which governs the coordinate systems in which those equations held.

---

19. "Zusammenfassend können wir sagen, dass die Nordströmsche Skalartheorie, welche an dem Postulat der Konstanz der Lichtgeschwindigkeit festhält, allen Bedingungen entspricht, die an eine Theorie der Gravitation beim heutigen Stande der Erfahrung gestellt werden können. Unbefriedigend bleibt leidiglich der Umstand, dass nach dieser Theorie die Trägheit der Körper beeinflusst, aber doch nicht verursacht erscheint, denn es ist nach dieser Theorie die Trägheit eines Körpers desto grösser, je weiter wir die übrigen Körper von ihm entfernt." Einstein, "Zum gegenwärtigen Stande...", p.1254.

20. A. Einstein, "Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie", Zeitschrift für Mathematik und Physik, 63 (1914), pp.215-25.

He concluded:

> "Now I am completely satisfied and no longer doubt the correctness of the whole system, whether the observation of the eclipse of the sun works out or not. The sense of the matter is too evident." [21]

In particular, he tells us elsewhere that he no longer felt that his 1913 field equations were inconsistent with the general relitivity of motion. In a 1914 discussion of the theory, he wrote on this question:

> "Conversely, physical processes must also determine the gravitational field, i.e. the quantities $g_{\mu\nu}$. The differential equations which determine these quantities were arrived at by means of the hypothesis that the laws of conservation of momentum and energy must be valid for material happenings and the gravitational field together. Consequently this hypothesis limits also the choice of the space-time variables x, without however, again arousing the epistemological considerations set out above. For, according to this generalised theory of relativity, physical qualities specific to preferred spaces no longer exist. The flow of all processes is governed by the quantities $g_{\mu\nu}$, which themselves are determined by the physical happenings of the whole of the remaining universe." [22]

---

21. "Nun bin ich vollkommen befriedigt und zweifle nicht mehr an der Richtigkeit des ganzen Systems, mag die Beobachtung der Sonnenfinsternis gelingen oder nicht. Die Vernunft der Sache ist zu evident." In P. Speziali, ed., Albert Einstein-Michele Besso: Correspondance 1903-1955 (Paris: Hermann, 1972), p.53.

22. "Umgekehrt müssen aber auch die physikalischen Vorgänge das Gravitationsfeld, d.h. die Grössen $g_{\mu\nu}$ bestimmen. Zu den diesen Grössen bestimmenden Differentialgleichungen gelangt man vermittelst der Hypothese, dass für das materielle Geschehen und das Gravitationsfeld zusammen die Erhaltungssätze des Impulses und der Energie gelten müssen. Diese Hypothese beschränkt nachträglich auch die Wahl der raum-zeitlichen Variabeln x, ohne dass jedoch dadurch die oben auseinandergesetzten erkenntnistheoretischen Bedenken wieder wachgerufen würden. Denn es gibt gemäss dieser verallgemeinerten Relativitätstheorie keine bevorzugten Räumen eigentümliche physikalische Qualitäten mehr. Der Ablauf aller Vorgänge wird durch die Grössen $g_{\mu\nu}$ beherrscht, die ihrerseits wieder durch das physikalische Geschehen des ganzen übrigen Weltalls bestimmt werden." A. Einstein, "Zum Relativitäts-problem," Scientia, 15(1914), p.348.

Elsewhere, Einstein dealt with this same question at greater length. [23]

From 1914, Einstein was satisfied that his theory was a true general theory of relativity, in spite of the fact that he was not to discover generally covariant field equations until the following year. From this time onwards, he presented the foundation and starting point of the theory as the recognition of the fact that both classical mechanics and his special theory of relativity suffered from a fundamental epistemological defect, the presence of inertial frames of reference as preferred frames of reference in space. From this he argued for the need for a theory which would realise the general relativity of motion and this entailed the relativity of inertia. Then, typically, through the principle of equivalence he would continue to argue that gravitation must play a role in the construction of such a theory. [24] These were the "epistemological considerations" referred to in the above quote.

This progression is almost a complete reversal of his development of the theory in his 1913 expositions. There he would present the theory initially as a theory of gravitation and, in the process of developing it, show how effects consistent with the relativity of inertia and the general relativity of motion could be recovered. [25]

---

23. A. Einstein, "Prinzipielles zur allgemeinerten Relativitätstheorie und Gravitations Theorie", Physikalische Zeitschrift, 15 (1914), pp.176-80.

24. See for example, Einstein, "Zum Relativitäts-problem", pp.344-8; A. Einstein, "Die formale Grundlage der allgemeinen Relativitätstheorie", Preussische Akademie der Wissenschaften, Sitzungsberichte, 1914 Part 2, pp.1030-2; Einstein, "The Foundation of...", pp.111-20.

25. See for example, Einstein, "Zum gegenwärtigen...", pp.1249-66; Einstein, "Physikalische Grundlagen...", pp.284-90.

The shifting of the foundation of the theory to the elimination of an epistemological defect in earlier theories is a most interesting step. It suggests that the key innovation of the theory lies not in the teasing out of the implications of a new empirical discovery, but that it is a result of pure speculative thought alone. I now review what Einstein saw this defect to be in the 1913 to 1916 period.

One of Einstein's earliest analyses of the question represented the defect as lying in a failure to recognise that motion, and acceleration in particular, only has meaning if it is defined with respect to other bodies. This led him directly to the relativity of inertia.

> "To speak of the motion, and therefore also the acceleration of a body A, in itself has no meaning. One can only speak of the motion and, correspondingly, the acceleration of a body A relative to other bodies B,C etc.. What holds in kinematical respect for the the acceleration should also hold for the inertial resistance, which the accelerated bodies oppose; it is to be expected a priori, even if not necessarily, that inertial resistance should be nothing other than a resistance against relative acceleration of the body A under consideration relative to the totality of all remaining bodies B,C etc.. It is well known that E. Mach in his history of mechanics first represented this standpoint with all sharpness and clarity, so that here I can refer simply to his exposition..." [26]

---

26. "Von Bewegung, also auch Beschleungigung eines Körpers A an sich zu reden, hat keinen Sinn. Man kann nur von Bewegung bzw. Beschleunigung eines Körpers A relativ zu anderen Körpern B,C usw. sprechen. Was in kinematischer Beziehung von der Beschleunigung gilt, das dürfte auch von dem Trägheitswiderstande gelten, den die Körper einer Beschleunigung entgegensetzen; es ist a priori zu erwarten, wenn auch nicht gerade notwendig, dass der Trägheitswiderstand nichts anderes sei als ein Widerstand gegen Relativbeschleunigung des betrachteten Körpers A gegenüber der Gesamtheit aller übrigen Körper B,C usw. Es ist wohlbekannt, dass E. Mach in seiner Geschichte der Mechanik diesen Standpunkt zuerst mit aller Schärfe und Klarheit vertreten hat, so dass ich hier einfach auf seine Ausführungen verweisen kann..." Einstein, "Zum gegenwärtigen...", p.1260.

To this passage, Einstein appended a footnote which contained an interesting extension of this argument.

> "One usually slips out of the consequences of considerations of this kind through the introduction of such reference systems in relation to which force free moving mass points execute a straightline uniform motion (Inertial systems). What is unsatisfactory about this is that it remains unexplained how inertial systems can be distinguished from other systems." 26a

These brief statements contain two distinct but related arguments. First Einstein asserts that the motion of a body only has meaning as motion with respect to other bodies and from this he concludes that inertial forces can only arise through acceleration with respect to other bodies. The second argument relies, in effect, on the principle of sufficient reason.[27] If, in contradiction to the general relativity of motion, we decide to admit inertial frames of reference as primitives in space, then, on the grounds of symmetry, there seems to be insufficient reason for one set of frames of reference to be distinguished as inertial rather than another. Further, in a later

---

26a. "Man pflegt den Konsequenzen derartiger Betrachtungen durch Einführung solcher Bezugssysteme zu entschlüpfen, in bezug auf welche kräftefrei bewegte Massenpunkte eine geradlinig gleichförmige Bewegung ausführen (Inertialsysteme). Das Unbefriedigende liegt dabei darin, dass unerklärt bleibt, wieso die Inertialsysteme gegenüber anderen Systemen ausgezeichnet sein können." Einstein, "Zum gegenwärtigen...", p.1260.

27. For discussion of this and other matters related to this chapter see J. Dorling, "Did Einstein need General Relativity to solve the Problem of Absolute Space? or had the Problem already been solved?", British Journal for the Philosophy of Science, 29 (1978), pp.311-23.

expansion of this argument, Einstein dismissed the introduction of inertial frames of reference as primitives as a purely ad hoc device for the explanation of the origin of inertial forces. Then he continued to conclude that the only admissible cause of certain frames of reference being inertial lies in the distant matter of the universe, in accord with the relativity of inertia. [28]

Einstein's other papers from this period contain no significant extension of the substance of these arguments. There is, however, careful elaboration and sometimes colourful illustration of them. For example, in a 1914 discussion, he presented a detailed dialogue, between someone untrained in traditional science and someone who is, over the explanation of the orbital motion of two bodies remote from the other matter of the universe. [29] The first notes the relationship between the bodies' motion and the fixed stars and is sure that there must be some connection. The second seeks to explain their motion in terms of the local inertial frames of reference, but, under pressure from the first, he is unable to give sufficient reason for why the particular frames that are inertial happened to be so.

In the paper, Einstein continued to a discussion based on the assertion that only relative acceleration has meaning. Then he turned to an illustration which here, at its first appearance, is attributed to his friend Besso. Consider an earlier time when the earth was still thought to be flat. The space would have been seen to admit a preferred direction - "downwards" - the direction in which bodies fall. The mistake here, Einstein tells us, is in attributing the

---

28. Einstein, "The Foundation...", pp.112-3.

29. Einstein, "Zum Relativitätsproblem", pp.344-6.

phenomenon of falling directly to an anisotropy in space, whereas its true cause lies in the presence of the earth. Similarly, we should not attribute the presence of inertial effects to the existence of preferred frames of reference in space. They should be explained in terms of the distribution of matter in the universe.

Einstein's 1916 presentation of these arguments is launched by a well known example. He considers two fluid spheres hovering in space. Each is observed to be rotating about the line joining them by observers at rest on the other. Only one, however, has an equatorial bulge. Einstein asks: "What is the reason for the difference in the two bodies?" and then proceeds as described earlier.[30]

If we analyse these arguments, we can see that they can be reduced in turn to a single hypothesis: the only causally efficacious entities in the universe are discrete bodies. Thus the only way we can judge the motion of a body is by reference to other bodies. In particular we cannot judge that a body is accelerating with respect to space from the presence of inertial forces. For, ex hypothesi, the only entities which are capable of causing such forces are the other bodies of the universe. Similarly, the hypothesis directly denies that a frame of reference may be inertial in itself and thus capable of causing inertial forces. It requires us to conclude that if acceleration with respect to a given frame of reference results in the appearance of inertial forces, then we must trace the origin of these forces to the masses present in some way.

---

30. Einstein, "The Foundation...", pp.112-3.

Now Einstein himself never reduced his analysis of the question in this period to the simple terms of this hypothesis. However, such a reduction is almost inescapable. For if we admit any other suitable entities as causally efficacious - for example Absolute space or inertial frames of reference - then Einstein's analysis no longer stands. Thus the establishment of Einstein's critique effectively becomes dependent on how one establishes this hypothesis.

In this period Einstein did not see the need to go into these questions beyond the depth of the considerations which we have seen discussed by him above. Rather he repeatedly attributed the origin of his analysis to Ernst Mach. Mach, as Einstein noted in the extract from his 1913 Vienna address quoted above, "...in his history of mechanics first represented this standpoint with all sharpness and clarity,.. so that here I can refer simply to his exposition...". In particular, Einstein attributed directly to Mach the idea that inertial forces arise through an interaction between all the bodies of the universe. In a letter written to Mach dated June 25th, 1913, in which he described some of the relevant results of his latest work on gravitation, he wrote:

> "For it necessarily turns out that inertia originates in a kind of interaction between bodies, quite in the sense of your considerations on Newton's pail experiment." [31]

In the next section, I follow Einstein's suggestion and refer to Mach's exposition of these questions and in particular his treatment of them in his history of mechanics. It was this exposition which Einstein repeatedly referred to in this 1912 - 1916 period. (See the above quote from his 1913 Vienna address and also his 1912 analysis,

---

31. Reproduced and translated in C.W. Misner, K.S. Thorne and J.A. Wheeler, Gravitation (San Francisco: Freeman, 1970), pp.544-5.

quoted in Section 4.5 earlier.)

6.3 Mach's Critique

Mach's critique of the Newtonian concepts of space and time has been the subject of much discussion for over a hundred years. In this section I will argue that Einstein and many others have been mistaken in their understanding of the conclusions that can be drawn validly from Mach's critique.[32] In particular, I will argue that the elimination of space as a causal entity, and the resulting conclusion that inertia must in some way arise as an interaction between a test mass and the other masses of the universe, do not follow from Mach's critique.[33]

According to Mach's sensationalist philosphy, the goal of science is the discovery of the simplest and most economical abstract summaries of sense experience.[34] Now for Mach sense experience consists of atomic sensations, such as "colors, sounds, temperatures, pressures, spaces, times, and so forth...".[35] These he preferred to

---

32. A similar view to the one presented here can be found in M. Strauss, "Einstein's Theories and the Critics of Newton", Synthese,18(1968), pp.251-84. See also J. Norton, "Einstein, Mach's Principle and the Origins of the General Theory of Relativity", Proceedings of the Australasian Association for the History, Philosophy and Social Studies of Science,12(1980-1), Forthcoming; M. Bunge, "Mach's Critique of Newtonian Mechanics", American Journal of Physics,34 (1966), pp.585-96; and H. Stein, "Some Philosophical Prehistory of General Relativity", Minnesota Studies in the Philosophy of Science, 8(1977), pp.3-49.

33. Many modern writers have recovered this result from Mach's critique. See for example, H. Reichenbach, The Philosophy of Space and Time (1927; New York: Dover, 1958), pp.213-7; D.W. Sciama, The Unity of the Universe (London: Faber & Faber, 1959), pp.84-102.

34. See E. Mach, The Analysis of Sensations (New York: Dover, 1959); E. Mach, "The Economical Nature of Physical Inquiry" pp.174-87 in J.J. Kockelmans, ed., The Philosophy of Science (New York: Free Press, 1968).

35. Mach, The Analysis..., p.2.

call "elements" in order to stress the fact that, on this level, he sought to make no distinction between subject and object. The familiar entities of experience, such as material bodies, and the ego were constructed as relatively stable complexes of these elements. Thus, in these terms, the goal of science, described above, was reduced effectively to the discovery of economical descriptions of the relations existing between these elements. In particular, he insisted on the elimination from these descriptions of all "metaphysical" entities, the "things-in-themselves", which were thought of as consisting of more than mere complexes of elements.

This orientation led Mach directly to his analysis of space and time. His sensationalism required him to reduce the then current understanding of space and time to economical descriptions of the relations between the elements of experience which comprise the relevant phenomena. In particular, Absolute Space and Absolute Time, in Newton's sense, are no longer to be admitted as anything more than a convenience in the discussion of these relations.

> "Since we only recognise what we call time and space by certain phenomena, spatial and temporal determinations are only determinations by means of other phenomena. If, for example, we express the positions of earthly bodies as functions of the time, that is to say, as functions of the earth's angle of rotation, we have simply determined the dependence of the positions of the earthly bodies on <u>one another</u>.
>
> "The earth's angle of rotation is very ready to our hand, and thus we easily substitute it for other phenomena which are connected with it but less accessible to us; it is a kind of money which we spend to avoid the inconvenient trading with phenomena, so that the proverb 'Time is money' has also a meaning here. We can eliminate time from every law of nature by putting in its place a phenomenon dependent on the earth's angle of rotation.

"The same holds of space. We know positions in space by the affection of our retina, of our optical or other measuring apparatus. And our x,y,z in the equations of physics are, indeed, nothing else than convenient names for these affects. Spatial determinations are, therefore, again determinations of phenomena by means of other phenomena.

"The present tendency of physics is to represent every phenomenon as a function of other phenomena and of certain spatial and temporal positions. If, now, we imagine the spatial and temporal positions replaced in the above manner, in the equations in question, we obtain simply _every phenomenon as a function of other phenomena_." [36]

Thus, when we say that some phenomenon stands in a given relationship with certain spaces or times, this is nothing more than a convenient summary for the equivalent statement that this phenomenon stands in the given relation to certain other phenomena which we represent in summary as the space or time. So if we wish to explicate descriptions of phenomena in which the terms "space" and "time" appear we only need to work through the relevant description replacing these terms by the appropriate phenomena. In his critique of Newton's concept of Absolute Space in Mach's _Science of Mechanics_, he explicates the concept of the motion of a body by the application of this method:

"When we say that a body K alters its direction and velocity solely through the influence of another body K', we have asserted a conception that it is impossible to come at unless other bodies A,B,C....are present with reference to which the motion of the body K has been estimated. In reality, therefore, we are simply cognizant of a relation of the body K to A,B,C...." [37]

---

36. E. Mach, _History and Root of the Principle of the Conservation of Energy_ (Chicago: Open Court, 1911), pp.61-2. Mach's emphasis.

37. E. Mach, _The Science of Mechanics_ (Lasalle, Ill.: Open Court, 1974), p.281.

It is from this persepctive that Mach then goes on to discuss the significance of Newton's bucket experiment. Newton had concluded that the effects appearing within the bucket had to be accounted for in terms of Absolute Space. This, of course, is unacceptable to Mach, for it involves the intrusion of a "thing-in-itself" into the economical descriptions of the relations between phenomena which comprise science. Rather, in a celebrated passage, he concludes the following:

> "Newton's experiment with the rotating vessel of water simply informs us, that the relative rotation of the water with respect to the sides of the vessel produce _no_ noticeable centrifugal forces, but that such forces _are_ produced by its relative rotation with respect to the mass of the earth and the other celestial bodies..." [38]

From the experiment, Mach feels entitled to conclude only a simple statement of sense experience on a relationship existing between certain phenomena - when there is relative rotation between the bucket and the earth and other celestial bodies, then there are centrifugal forces within the bucket. In particular, he is _not_ postulating an underlying mechanism which would cause these forces, which, for example, could be extrapolated to other cases. So he continues:

> "...No one is competent to say how the experiment would turn out if the sides of the vessel increased in thickness and mass till they were ultimately several leagues thick..."

This sentence has been misinterpreted repeatedly and taken to suggest that if the vessel were increased in size and mass in the way described, then its rotation could also induce centrifugal forces inside the

---

38. _Ibid._, p.284. Mach's emphasis.

bucket.[39] This is clearly not what Mach intended, as is clear from the paragraph preceding this passage and the sentence which immediately follows:

> "...The one experiment only lies before us, and our business is, to bring it into accord with the other facts known to us, and not with the arbitrary fictions of our imagination."

For Mach the point of this example is to demonstrate the effectiveness of his sensationalism as a means of avoiding idle metaphysical speculation slipping into our discourse. Provided we carefully limit ourselves to the description of the phenomena given by experience, then we have no need of Absolute Space. This need only arises if we go beyond the phenomena given by experience and consider, for example, the rotation of the earth or the motion of a body K in the absence of the other bodies of the universe. Newton's bucket experiment tells us nothing about such cases, nor does it justify the introduction of Absolute Space. Within Mach's sensationalist framework, it tells us only that centrifugal forces arise within the bucket when there is relative rotation between it and the earth and other bodies of the universe.

Now Einstein's claim that the relativity of inertia follows from Newton's bucket experiment involves exactly this type of illegitimate extrapolation from the phenomena given by experience. For the relativity of inertia involves the assumption that inertial forces will act on a test body surrounded by a rotating shell of matter, considered as a separate, isolated case. Such a case was not given by experience to Einstein. Einstein must sum over an

---

39. Reichenbach, p.214.

infinite number of such cases, combining an infinite number of shells to represent all the matter of the universe, just to recover the single case given by experience, that of Newton's bucket. Thus the requirements of Mach's critique rules out the relativity of inertia just as much as it does Absolute Space.

Further, Mach's critique did not require the elimination of all entities excepting discrete bodies as causally efficacious. Rather, it involved a far stronger requirement, the elimination of all talk of causal processes whatsoever, in the sense of Einstein's later usage of the term as relating to underlying mechanisms. Continuing the passage quoted earlier from his History and Root of the Principle of the Conservation of Energy, Mach wrote:

> "Thus the law of causality is sufficiently characterized by saying that it is the presupposition of the mutual dependence of phenomena. Certain idle questions, for example, whether the cause precedes or is simultaneous with the effect, then vanish by themselves.
>
> "The law of causality is identical with the supposition that between natural phenomena α, β, γ, δ, ... ω certain equations subsist..." [40]

This directly contradicts the account of causation implicit in Einstein's treatment of the relativity of inertia. In that account, we unambiguously locate the cause of inertial forces induced in a test mass in the rotating shell of matter surrounding it and, since action from the shell propagates with finite speed, we must conclude that this cause precedes the effect.

In conclusion, we can say that Mach proposed an explication of the concepts of space and time in Newtonian theory by a rewriting

---

40. Mach, History and Root..., p.62. Mach's emphasis.

of the essential contents of that theory as an economical catalogue of sense experience. This involved the elimination of the terms "space" and "time" from the relevant discourses and their replacement by phenomena accessible to sense experience. Einstein, however, understood the critique as demanding a revision of the content rather than the presentation of Newtonian theory. This would involve the elimination of space, in particular, as a causal entity and the construction of a new mechanics which could locate the source of inertial forces in the distant matter of the universe.

In short, Einstein understood the critique to require that all kinematically equivalent systems be also dynamically equivalent. Thus, since all states of acceleration are kinematically equivalent, he believed that this equivalence should extend also to the inertial forces which arise in association with them. However, in this regard, all the critique really required was that one replace dynamical accounts by purely phenomenal, kinematical accounts in order to eliminate talk of the metaphysical entities in them.

6.4 The Elimination of Space

In the last section, I argued that Einstein could not claim support for his concepts of the general relativity of motion and the relativity of inertia from Mach's critique of Newtonian mechanics. However, this is by no means equivalent to the dismissal of these concepts. So in this section I turn to analyse their viability in the relevant context.

In many circles, these concepts have become an accepted feature of modern physical theory since the early decades of this century.

The relativity of inertia, in particular, has been invoked often by eminent writers in a wide range of contexts, usally bearing the name "Mach's Principle". Although this name was first used by Einstein, we shall see that he did not use it to refer to the relativity of inertia, contrary to the practice which is now widespread. There is still a certain amount of ambiguity in exactly what the term does refer to in modern usage. Pirani has been able to distinguish five distinct meanings. [40a]

Mach's principle, as the relativity of inertia, has been invoked frequently as a part of the structure of the general theory of relativity. [41] It has also been invoked outside of the context of this theory. In particular, attempts to construct a properly "Machian" mechanics are frequent and it has also been used in a weaker sense to justify modifications to the general theory of relativity. [42]

40a F.A.E. Pirani, "On the Definition of Inertial Systems in General Relativity", Helvetica Physica Acta, 1956 Suppl.IV, pp.198-203. The literature in this general area is so expansive that it would be impossible and inappropriate to give it anymore than a fairly superficial glance in this section. See M. Reinhardt, "Mach's Principle - A Critical Review", Zeitschr. Naturforsch.,28 (1973), pp.529-37.

41. See for example M. Born, Einstein's Theory of Relativity (New York: Dover, 1962), pp.308-12; Reichenbach, pp.213-7; W. Davidson, "General Relativity and Mach's Principle", Monthly Notices of the Royal Astronomical Society,117(1957), pp.212-24.

42. See for example D.W. Sciama, "On the Origin of Inertia", Monthly Notices of the Royal Astronomical Society,113(1953), pp.34-42; R.H. Dicke, "Cosmology, Mach's Principle and Relativity", American Journal of Physics,31(1963), pp.500-9; R. Goldoni, "A Background-Dependent Approach to the Theory of Gravitation", General Relativity and Gravitation,7(1976), pp.731-41, 743-55; H.J. Treder "Telescopic Principles in the Theory of Gravitation...", Annalen der Physik,36(1979), pp.4-19. See also V.W. Hughes, "Mach's Principle and Experiments on Mass Anisotropy" in H.Y. Chiu and W.F. Hoffmann, eds., Gravitation and Relativity (New York: Benjamin, 1964), pp.106-20; P.W. Bridgman, "Significance of the Mach Principle", American Journal of Physics,29(1966), pp.32-36.

In general such attempts do not contain serious analysis of the justification for the principle itself. Nevertheless, in this section, I seek to establish that the relativity of inertia and Einstein's early concept of the general relativity of the motion of bodies are not a part of the theory, as it stood formally in 1916, [43] and, moreover, that they are inconsistent with the world view promoted by the theory. [44]

The first point can be established readily by noting that this form of the theory admits many possible universes as solutions of the field equations which are inconsistent with the above results. Consider, for example, the flat Lorentzian spacetime of a matter free universe. In such a universe there is a preferred set of frames of reference, the set of inertial frames of reference, but there is certainly no sufficient reason provided for why these particular ones should be inertial, in accordance with the demands of Einstein's analysis. Also, the inertia of a test body cannot be regarded as being determined by the other masses of the universe, for there are none. Similar conclusions follow from the consideration of the spacetime which contains a lone island of matter and which asymptotically approaches a Lorentzian structure, with a preselected set of inertial frames at spatial infinity. Fock is well known for

---

43. As in Einstein, "The Foundation of...".

44. See also for questions relating to this section and the remainder of the chapter: J.C. Graves, The Conceptual Foundations of Contemporary Relativity Theory (Cambridge, Mass.: M.I.T. Press, 1971), pp.289-305; L. Sklar, Space, Time and Spacetime (Berkeley: University of California Press, 1977), pp.210-24; M.R. Gardner, "Relationism and Relativity", British Journal for the Philosophy of Science, 28 (1977), pp.215-33.

his advocacy of the position that Einstein's theory is not a theory of general relativity at all, on the basis of such considerations.[45]

In maintaining such a position, it is important not to overlook the fact that, whilst Einstein's theory may not be a true theory of general relativity, it certainly has made substantial steps towards one. Unlike the theories which have gone before it, Einstein's theory no longer needs any presuppositions about the disposition of certain preferred frames of reference in space, as does the special theory of relativity, for example, with its set of inertial frames of reference. This is not to say that Einstein's general theory makes no presuppositions about the structure of spacetime. Rather, the theory allows its general structure to be more negotiable than ever before. In the special theory of relativity, one needs to stipulate as an external and prior condition, which frames of reference are inertial. As we shall see, Einstein's general theory leads to the possibility of universes in which the disposition of inertial frames of reference is determined entirely by the distribution of matter. This type of negotiability of the geometry of spacetime has been described by the compact term "no prior geometry".[46]

The theory can have this property because it is essentially a local theory, which deals with the relationship between the structure of spacetime and matter in infinitesimally small regions, whereas such questions as the existence of preferred frames of reference

---

45. V. Fock, The Theory of Space, Time and Gravitation (London: Pergamon, 1959).

46. C.W. Misner, K.S. Thorne and J.A. Wheeler, Gravitation (San Francisco: Freeman, 1970), pp.429-31.

usually arise through global considerations. This provides a way to escape the theory's failure to contain the general relativity of motion and the relativity of inertia. We simply add a global constraint to the theory which forbids such solutions of the field equations as the ones described above. We shall see that Einstein adopted a similar approach later and that related approaches have been used by others more recently.[47]

Provided we can agree on how these requirements are to be reformulated, we can retain them in the context of the general theory of relativity. I will argue, however, that such a tactic acts only to mask more fundamental inconsistencies between them and the world view promoted by the theory, as it is now understood. I stress that this does not necessarily reflect on the reformulations themselves, for we are free to consider them on their own individual merits. It affects them only in so far as we seek to justify them on the basis of Einstein's original critique.

Earlier I argued that Einstein's critique was based on the assumption that the only causally efficacious entities in the universe were discrete bodies and the success of his analysis depends on the extent to which this assumption could be established. Now Mach's original critique and Einstein's later interpretation of it were made within the corpuscular world view promoted by Newtonian mechanics. In this world view, the universe is seen to consist of discrete masses in motion within an essentially featureless container of space. From

---

47. For example, J.A. Wheeler, "Mach's Principle as a Boundary Condition for Einstein's Field Equations" in W.E. Brittin, B.W. Downs and J. Downs, eds., Lectures in Theoretical Physics Vol.V (New York: Interscience, 1963), pp.528-85.

here it is but a small step to remove the very last active power from space and locate the source of inertial forces in the masses of the universe. Thus Einstein's early concept of the general relativity of motion and the relativity of inertia arise naturally within the world view promoted by Newtonian mechanics. In particular, the appearance of inertial forces as the result of an interaction between an accelerated test mass and the other distant masses of the universe can be understood most naturally as a typical action at a distance effect.

This world view, however, breaks down when we shift to the conceptual framework of relativity theory. A fundamental assumption of the theory is the assumption that there is a maximum speed at which effects can propagate through space and therefore that the interactions of separated bodies must be mediated by the action of fields. These fields come to be seen to have an independent existence of their own. Thus, far from being the only causally efficacious entities, bodies find themselves surrounded by a space permeated with causally efficacious fields.

When we come to seek the origins of the inertial forces acting on a body, this new approach directs us to look in the immediate neighbourhood of the body and to expect these forces to arise through an interaction between the body and some suitable field. Whatever influence the distant masses of the universe may have in this process would have to be somewhat indirect, for such an influence would need to be carried by a field and could only propagate at a certain finite speed. But since we can no longer assume that these distant masses are the only causally efficacious entities in the universe, it is not immediately clear why we should expect them to play a role at all. Einstein originally traced the source of inertial forces to these masses by a process of elimination. They were the only causal entities available for the task. Clearly, we can no longer do the same.

Within the general theory, the field with which bodies interact to produce inertial and gravitational forces is identified as the metric field of spacetime itself. In this theory the notion of acceleration with respect to space, as opposed to other bodies, does not become entirely meaningless as Einstein originally suggested. For, within the theory, certain states of motion, such as acceleration and uniform motion, are distinguished by the nature of the local interaction between the body and the metric field. Indeed, if the mass of the moving body is not vanishingly small, then the interaction can become quite a dynamic one, with the energy and momentum of the external metric field exchanging with that of the body itself.

Within the theory, fields are seen to carry energy and momentum, just as normal matter does, and they are recognised as having an equivalent ontic status. As we shall see, the theory even led to the view that fields may be more fundamental than normal matter itself and that particles are to be reduced, for example, to singularities in these fields. Within such a conception, the general relativity of the motion of bodies is far from being a fundamental postulate of the theory and, if at all, would be expected to arise only as a theorem late in the development of the theory.

Further, Einstein originally understood his new theory to amount to the final elimination of space and time. Yet it turned out to lead to quite a contrary view. Special relativity had already blurred the distinction between matter and fields and now Einstein's general theory, in fusing gravitational fields with the structure of space-time itself, had begun to blur the distinction between matter, the contained, and spacetime, the container. Rather than eliminating space and time, Einstein's theory suggested their elevation by incorporation into a new entity which would embrace all matter, fields and spacetime, a unified field, the new Unity of the universe.

Thus we can conclude that the general theory of relativity ultimately led to a world view which contradicted the fundamental concepts of the critique which Einstein believed had given birth to it. In the sections that follow, I examine how Einstein came to develop this new world view out of his theory, in the years following 1916, and how he reassessed and modified the essential concepts of his original critique.

6.5 Einstein and the Status of Fields

Einstein's early analyses of the relativity of motion and the origin of inertia were carried out within a conceptual framework promoted by Newtonian mechanics, in which the universe is seen to consist of discrete moving bodies in the essentially featureless container of space. This world view was displaced through the emergence of the field concept in the nineteenth century and its victory was finally assured with the establishment of the special theory of relativity.

However, even as late as 1915, Einstein had neither explicitly acknowledged the new status of fields through his publications nor implicitly in his critiques of the nature of space, time and motion. In 1916 and the years following, this began to change. Early in this process, in his popular exposition of the special and general theories of relativity, he introduced a brief discussion of the field concept, but only felt the need for it when he came to discuss the general theory and the nature of gravitation. There he wrote as follows:

> "As a result of the more careful study of electromagnetic phenomena, we have come to regard action at a distance as a process impossible without the intervention of some intermediary medium. If, for instance, a magnet attracts a piece of iron, we cannot be content to regard this as meaning that the magnet acts directly on the iron through the intermediate empty space, but we are constrained to imagine - after the manner of Faraday - that the magnet always calls into being something physically real in the space around it, that something being what we call a 'magnetic field'. In its turn this magnetic field operates on the piece of iron, so that the latter strives to move towards the magnet. We shall not discuss here the justification for this incidental conception, which is indeed a somewhat arbitrary one. We shall only mention that with its aid electromagnetic phenomena can be theoretically represented much more satisfactorily than without it, and this applies particularly to the transmission of electromagnetic waves. The effects of gravitation are also regarded in an analogous manner." [48]

The latter part of this passage reflects Einstein's early positivist and in this case instrumentalist approach to fields. They were merely an "incidental conception" and a "somewhat arbitrary one" at that. But they did enable what he then saw as really important, the actual phenomena themselves, to be "represented much more satisfactorily", as is the role of any theoretical term in an instrumentalist account. Their ontic status, however, is open to

---

48. A. Einstein, Relativity. The Special and the General Theory (London: Methuen, 1977), pp.63-4.

question since representation of the relevant phenomena is quite conceivable without them.

It was within such a conceptual framework that Einstein, earlier in 1915, could commend his new generally covariant field equations to us "...through which space and time are robbed of the last trace of objective reality..." [49] and conclude that his requirement of general covariance "...takes away from space the last remnant of physical objectivity...".[50]

The passage quoted above is of special interest, because it displays clearly the transition in progress in Einstein's thought in the 1916-1917 period. As well as containing a compact statement of Einstein's earlier instrumentalist approach to the field concept, we also see the first suggestions of the realist view of fields, which was to dominate Einstein's work for the remainder of his life. For, we are "constrained to imagine" the entity which mediates magnetic interactions between separated bodies as "physically real".

In the years immediately following this period Einstein had occasion to reflect on and reassess the foundations of his new theory and the critiques that supported it. During this time, it is known that Einstein began to feel a disenchantment with Mach's philosophical system and, this may have influenced the direction of his thoughts. [51]

---

49. "...durch welche Zeit und Raum der letzten Spur objektiver Realität beraubt werden..." A. Einstein, "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie", <u>Preussische Akademie der Wissenschaften, Sitzungsberichte</u>, 1915 Part 2, p.831.

50. Einstein, "The Foundation of...", p.117.

51. G. Holton, "Mach, Einstein and the Search for Reality", <u>Daedalus</u>, <u>97</u>(1968), pp.636-73.

By 1920, Einstein was no longer to insist that his theory of relativity had eliminated the ether from physical theory. He wrote:

> "More careful reflection teaches us, however, that the special theory of relativity does not compel us to deny ether. We may assume the existence of an ether; only we must give up ascribing a definite state of motion to it, i.e. we must by abstraction take from it the last mechanical characteristic which Lorentz had still left it. We shall see later that this point of view...is justified by the results of the general theory of relativity." [52]

From this, Einstein went on to sketch the place of what he now called the "ether" in the development of physical theory from Newton to Mach and Lorentz and on to his general theory of relativity. The reintroduction of the term "ether" comes at a time when the status of the ether in physical theory was under serious attack, as a direct result of Einstein's own writings on the theory of relativity. However, in Einstein's hands the use of the term does not amount to a reversion to the older concepts of the ether as he went to pains to explain, rather it amounts to a recognition of the fact that his theory had by no means eliminated space as an entity from physical theory.

In a later 1924 paper, "On the Ether", he expanded on this. [53] In introducing the paper, he explained that one could take the term "physical qualities of space" equally well in place of "ether" [54] and then, under the banner of giving a history of the development of the ether concept, gives what we would now understand to be a history of the development of the concepts of space and time, as they lead

---

52. A. Einstein, "Ether and Relativity" in Sidelights on Relativity (London: Methuen, 1922), p.13.

53. A. Einstein, "Über den Äther", Schweizerische Naturforschende Gesellschaft, Verhandlungen, 105 (1924), pp.85-93.

54. "physikalischen Qualitäten des Raumes" Ibid., p.85.

up to his theory of relativity. This account follows the general plan of his later accounts of the development of these concepts, although he preferred later to suppress the term "ether" in this context.[55]

At the conclusion of this 1924 paper, Einstein related his resurrection of the ether and space to the field concept and the results of his theory of relativity. After a brief discussion of recent developments in quantum theory, he wrote:

> "But even if these possibilities mature to real theories, in theoretical physics we shall not be able to dispense with the ether, i.e. with a continuum provided with physical properties; for the general theory of relativity, on the points of view of whose principles physicists will probably always hold fast, excludes unmediated action at a distance; each local action theory however, assumes continuous fields, therefore also the existence of an 'ether'".[56]

In the course of these years following Einstein's discovery of his general theory of relativity, his increasing commitment to the concept of the field and the retention of space was directed into his almost legendary search for a unified field theory. This theory

---

55. See for example A. Einstein, Ideas and Opinions, (London: Souvenir 1973).

56. "Aber selbst wenn dies Möglichkeiten zu wirklichen Theorien heranreifen, werden wir des Äethers, d.h. des mit physikalischen Eigenschaften ausgestatteten Kontinuums, in der theoretischen Physik nicht entbehren können; denn die allgemeine Relativitätstheorie, an deren grundsätzlichen Gesichtpunkten die Physiker wohl stets festhalten werden, schliesst eine unvermittelte Fernwirkung aus; jede Nahewirkungs-Theorie aber setzt kontinuierliche Felder voraus, also auch die Existenz eines 'Äthers'". Einstein, "Über den Äther", p.93.

was to be the culmination of the emergence of the field concept and his theory of relativity. It postulated the existence of a single field in which all known entities, the metric field, the gravitational field, the electromagnetic field and all forms of matter, were to be combined. Far from robbing space of the last trace of physical objectivity, he expected it to elevate space to a status of unique importance. I quote a 1930 statement of this:

> "...the axiomatic foundation of physics appears as follows. The real is conceived as a four-dimensional continuum with a unitary structure of a definite kind (metric and direction). The laws are differential equations, which the structure mentioned satisfies, namely, the fields which appear as gravitation and electromagnetism. The material particles are positions of high density without singularity.
>
> "We may summarize in symbolical language. Space, brought to light by the corporeal object, made a physical reality by Newton, has in the last few decades swallowed ether and time and seems about to swallow also the field and the corpuscles, so that it remains as the sole medium of reality." [57]

This complete reversal in Einstein's understanding of the significance of his theory of relativity is dramatic to say the least. Rather than banishing space and time from the physical world, his theory set him on a path which he hoped would lead to the establishment of space as the "sole medium of reality". Running parallel to this reversal came an explicit acknowledgement of the change in world view ushered in by the emergence of the field concept. In a 1931 account of Maxwell's role in this change, he described the earlier view in the following way:

57. A. Einstein, "Space, Ether and the Field in Physics", Forum Philosophicum, 1 (1930), p.184.

> "According to Newton's system, physical reality is characterized by the concepts of space, time, material point, and force (reciprocal action of material points). Physical events, in Newton's view, are to be regarded as the motions, governed by fixed laws, of material points in space. The material point is our only mode of representing reality when dealing with changes taking place in it, the solitary representative of the real, in so far as the real is capable of change." [58]

Elsewhere Einstein coupled this view with the concept of action at a distance. [59] But here he went on to describe the impact of the emergence of the field concept:

> "...before Maxwell people conceived of physical reality - in so far as it is supposed to represent events in nature - as material points, whose changes consist exclusively of motions, which are subject to total differential equations. After Maxwell they conceived physical reality as represented by continuous fields, not mechanically explicable, which are subject to partial differential equations. This change in the conception of reality is the most profound and fruitful one that has come to physics since Newton..." [60]

We shall see in the next section that Einstein came to recognise that his concept of the relativity of inertia belonged essentially to the earlier view of reality. But he was to retain the belief that the critique which supported this concept has touched upon the crucial weakness of classical mechanics and his special theory of relativity and that his general theory embodies the removal of this defect.

The concept of the field remained at the heart of Einstein's work for the rest of his life. In 1954, Einstein wrote his last letter to his life-long friend Michele Besso. In the latter he

---

58. A. Einstein, "Maxwell's Influence on the Evolution of the Idea of Physical Reality", in Ideas and Opinions (London: Souvenir, 1973), pp.266-7.

59. A. Einstein, "The Fundaments of Theoretical Physics", in Ideas and Opinions (London:Souvenir, 1973), p.325.

60. Einstein, "Maxwell's Influence...",p.269.

summarised the achievements of his theory of relativity and his unified field theory, work which had spanned the previous fifty years. He concluded:

> "I concede, however, that it is quite possible that physics cannot be founded on the concept of the field - that is to say, on continuous elements. But then, out of my whole castle in the air - including the theory of gravitation, but also most of current physics - there would remain almost <u>nothing</u>." [61]

<u>6.6 "Mach's Principle"</u>

The hypothesis of the relativity of inertia had played an important role in Einstein's discovery of the general theory of relativity. But, in 1916, its status within the theory was not completely clear. In his well known 1916 exposition of the theory, Einstein used this hypothesis to justify his attempts to construct a general theory of relativity, but he left open the question of whether it was actually contained within the theory. [62]

In the years immediately following, Einstein sought to establish the relativity of inertia through his new theory and in so doing came to face the difficulties discussed in the preceding sections. First he sought to invoke the hypothesis as a boundary condition on solutions of the field equations. Then, as his awareness of the importance of the field concept to his theory grew, he was led to seek more radical reformulations of the hypothesis and eventually he he was driven to the conclusion that essential aspects of his original

---

61. Quoted from the translation in P. Speziali, "Einstein Writes to his Best Friend" in A.P. French, ed., <u>Einstein. A Centenary Survey</u> (Cambridge, Mass.: Harvard Univ. Press,1979), p.269. Einstein's emphasis.

62. Einstein, "The Foundation of...".

idea were incompatible with his general theory of relativity.

Einstein's first steps on this path were reported by de Sitter in a 1916 paper.[63] There he described how the usual assumption that the metric should reduce to a simple Lorentzian form in regions remote from matter was inconsistent with Einstein's ideas on the origin of inertia. He went on to report that Einstein had suggested to him in a conversation that, in such regions, the metric might reduce to the form

$$\begin{bmatrix} 0 & 0 & 0 & \infty \\ 0 & 0 & 0 & \infty \\ 0 & 0 & 0 & \infty \\ \infty & \infty & \infty & \infty^2 \end{bmatrix}$$

where the non-zero terms are the time components. This degenerate form of the metric had suitable transformation properties and predicted that a test body remote from all other bodies would have no inertia.

In 1917, Einstein published on the question.[64] He promised to "...conduct the reader over the road that I myself have travelled..." and went on to describe how he had tried to incorporate the relativity of inertia into his theory. He wrote:

> "The opinion which I entertained until recently, as to the limiting conditions to be laid down in spatial infinity, took its stand on the following considerations. In a consistent theory of relativity there can be no inertia relatively to 'space', but only an inertia of masses relatively to one another. If, therefore, I have a mass at a sufficient distance from all other masses in the universe, its inertia must fall to zero. We will try to formulate this condition mathematically."[65]

---

63. W. de Sitter, "On Einstein's Theory of Gravitation and its Astronomical Contents", Monthly Notices of the Royal Astronomical Society, 76 (Suppl.1916), pp.699-728; 77(1916), pp.155-84; 78 (1917), p.3-28. See 77(1916), pp.181-2.

64. A. Einstein, "Cosmological Considerations on the General Theory of Relativity" in The Principle of Relativity (New York: Dover, 1952), pp.175-88.

65. Ibid., p.180. Einstein's emphasis.

He proceeded, by means of a simplified case, to argue for the degeneracy of the metric, as described above, in regions remote from bodies and then to describe how it had turned out to be impossible to impose this degeneracy as a boundary condition on suitable solutions of the field equations.

This last result posed a very serious threat to the idea that Einstein's theory was consistent with the relativity of inertia, let alone that it contained the hypothesis. For now it seemed that the structures of the universes predicted by the theory were inconsistent with the hypothesis. However, Einstein was firmly wedded to the idea that his theory enabled the relativity of inertia to be realised. He proposed an ingenious way out. He noted that the problem only arose when one tried to impose boundary conditions at infinity for particular solutions of the field equations. What if the real universe were spatially closed and thus spatially finite? Then one would have no need of such boundary conditions and the problem would no longer arise.

This solution was far from satisfactory from the very start. In order to be able to recover a closed static universe from his theory, he found it necessary to supplement his 1915 field equations with an extra ad hoc term, which involved the introduction of the controversial cosmological constant. Within two years he was to describe this addition as "gravely detrimental to the formal beauty of the theory".[66] With the discovery of the expansion

---

66. A. Einstein, "Do Gravitational Fields Play an Essential Part in the Structure of the Elementary Particles of Matter" in The Principle of Relativity (New York: Dover, 1952), p.193.

of the universe at the end of the 1920's, the superfluity of the term became very clear. Had Einstein insisted on closure of the universe without the additional terms, he might have been able to predict the expansion. It is well known that Einstein gave up its use [67] and was later to describe its introduction to Gamow as "...one of the biggest blunders he had made in his entire life". [68]

Further, Einstein had to face the arguments of de Sitter who had remained in close contact with Einstein over this question. De Sitter had maintained a persistent and pertinent critique of Einstein's analysis of the origin of inertia. [69] He noted that if inertia were caused by the other bodies of the universe, then it was certain that the matter of the universe known to astronomers played no significant role in this process. Thus, de Sitter argued, Einstein was forced to attribute the origin of inertia to masses beyond the realm of observation, a conclusion which was as objectionable as Newton's introduction of Absolute Space. De Sitter was also able to show that there were other solutions to Einstein's augmented field equations

---

67. See for example A. Einstein, Review of R. Tolman: Relativity, Thermodynamics and Cosmology, Science, 80 (1935), p.358.

68. G. Gamow, "The Evolutionary Universe" in The Universe (London: Bell and sons, 1958), p.67.

69. See de Sitter, op.cit.; also W. de Sitter, "On the Relativity of Rotation in Einstein's Theory", Proceedings of the Academy of Science, Amsterdam, 19 (1917), pp.527-32; W. de Sitter, "On the Relativity of Inertia. Remarks on Einstein's Latest Hypothesis", Proceedings of the Academy of Science Amsterdam, 19 (1917), pp.1217-1225; W. de Sitter, "On the Curvature of Space", Proceedings of the Academy of Science, Amsterdam, 20 (1917), pp.229-42; W. de Sitter, "Further Remarks on the Solutions of the Field Equations of Einstein's Theory of Gravitation", Proceedings of the Academy of Science, Amsterdam, 20 (1918), pp.1309-12.

and that these solutions were matter free. This was clearly inconsistent with Einstein's concept of the origin of inertia, for in such universes the inertia of an infinitesimal test body could not be attributed to the other bodies of the universe.

In his 1917 paper, Einstein introduced the closed universe because it enabled him to avoid the need for boundary conditions inconsistent with his hypothesis of the relativity of inertia. However, he did not explain why he felt such a universe captured the key features of his understanding of the origin of inertia. This was explained in a 1918 paper, in which he introduced a principle for which he chose the name "Mach's principle". It read as follows:

> "Mach's Principle: The G-field [metric field] is determined completely through the masses of bodies. Since, following the results of the special theory of relativity, mass and energy are the same and energy is formally described through the symmetrical energy tensor ($T_{\mu\nu}$), this means that the G-field should be conditioned and determined through the energy tensor of matter."

To this he appended an illuminating footnote:

> "So far I have not distinguished principles (a) [the principle of relativity] and (c) [Mach's principle], which caused confusion. Therefore I have chosen the name Mach's principle, because this principle signifies a generalisation of Mach's requirement that inertia must be reduced to an interaction of bodies." [70]

---

70. "Machsches Prinzip: Das G-Feld ist restlos durch die Massen der Körper bestimmt. Da Masse und Energie Nach den Ergebnissen der speziellen Relativitätstheorie das Gleiche sind und die Energie formal durch den symmetrischen Energietensor ($T_{\mu\nu}$) beschreiben wird, so besagt dies, dass das G-Feld durch den Energietensor der Materie bedingt und bestimmt sei."

"Bisher habe ich die Prinzipe (a) und (c) nicht auseinandergehalten, was aber verwirrend wirkte. Den Namen 'Machsches Prinzip' habe ich deshalb gewählt, weil dies Prinzip eine Verallgemeinerung der Machschen Forderung bedeutet, dass die Trägheit auf eine Wechselwirkung der Körper zurückgeführt werden müsse." A. Einstein, "Prinzipielles zur Allgemeinen Relativitätstheorie", Annalen der Physik, 55 (1918), pp.241-2. Einstein's emphasis.

This Mach's principle was by no means a simple restatement of the content of the hypothesis of the relativity of inertia. This latter hypothesis was defined within the corpuscular view of the cosmos and dealt with the interactions of discrete bodies only. The new principle, however, dealt with entirely different entities, the continuous fields of the general theory of relativity, whose status in the theory was only then coming to be recognised by Einstein. It is clear that the relativity of inertia inspired this statement of Mach's principle. However, as I have argued earlier, the arguments which supported the relativity of inertia were based on an assumption that was untenable in the field theoretic framework in which the principle was presented. Thus some justification for the new principle would not be inappropriate. However, Einstein seemed to be satisfied to justify the principle as an extension of his earlier hypothesis of the relativity of inertia, which, in the relevant footnote, is attributed by him to Mach. In any case, he went on to tell us that, in spite of the doubts of some of his colleagues, he found the satisfaction of this principle to be "unconditionally necessary".[71]

Einstein then continued to relate his new principle to his 1917 paper. In particular he noted that there appears to be no singularity free spacetime admitted by the augmented field equations in the case when the energy tensor of matter disappears everywhere. This of course is not the case for his original 1915 equations. This suggests

---

71. "unbedingt notwendig". Ibid.,p.242.

that matter is the sole source of spacetime and that without matter there can be no spacetime. That is, without matter an infinitestimal test body can have no inertia, for there is no spacetime structure with which to interact.

The existence of de Sitter's matter free solution to the new field equations directly contradicted this view. Einstein sought to dismiss this solution as containing physically unacceptable singularities.[72] Later writers regard de Sitter's result as a serious blow to Einstein's position. However, at the time, Einstein appeared to be untroubled by it.[73] He continued to assert for the origin of inertia if the universe were spatially closed. Moreoever at this time he took his 1918 Mach's principle to be a natural implementation of what he understood Mach's critique of the origin of inertial forces to be and linked the principle closely with the relativity of inertia.[74]

---

72. A. Einstein, "Kritisches zu einer von Hrn. de Sitter gegebenen Lösungen der Gravitationsgleichungen", Preussische Akademie der Wissenschaften, Sitzungsberichte, 1918, part 1, pp.270-2. Then see de Sitter, "Further Remarks..." op.cit.

73. J.D. North, The Measure of the Unvierse (Oxford: Clarendon, 1965), pp.87-92; M. Jammer, Concepts of Space (Cambridge, Mass.: Harvard U.P., 1969), pp.191-9.

74. See A. Einstein, "Geometry and Experience" in Ideas and Opinions (London: Souvenir, 1973), p.239; A. Einstein, "On the Theory of Relativity" in Ideas and Opinions (London: Souvenir, 1973, pp.248-9; A. Einstein, "Fundamental Ideas and Problems of the Theory of Relativity" in G.E. Tauber, ed., Albert Einstein's Theory of General Relativity (New York: Crown,1979), pp.56-7; Einstein, The Meaning of Relativity, pp.95-103.

After this Einstein quietly dropped the idea that the origin of inertia could only be satisfactorily dealt with in his theory if the universe were closed. It is not clear exactly what led him to do this. Presumably it stemmed from his concern that the cosmological term was gravely detrimental to the beauty of his theory and his rejection of the term following the discovery of the expansion of the universe. His mature view is presented in this Autobiographical Notes as follows:

> "Mach conjectures that in a truly reasonable theory inertia would have to depend upon the interaction of the masses, precisely as was true for Newton's other forces, a conception that for a long time I considered in principle the correct one. It presupposes implicitly, however, that the basic theory should be of the general type of Newton's mechanics: masses and their interactions as the original concepts. Such an attempt at a resolution does not fit into a consistent field theory, as will be immediately recognized." [75]

It is interesting to note that the essential feature of this argument, the idea that the relativity of inertia belongs to the realm of Newtonian action at a distance theory and not the realm of the field theory of general relativity, was already presented by Einstein as early as 1920. [76]

Thus Einstein came to recognise the incompatability of his hypothesis of the relativity of inertia and the fundamental concepts of his general theory of relativity. Further, his attempts to salvage the essential content of the hypothesis were not successful and faded out of his accounts of the theory. But Einstein never lost his conviction that Mach's critique had uncovered a fundamental defect of Newtonian mechanics, which was shared by his special theory of relativity, and that his general theory had succeeded in eliminating it.

---

75. A. Einstein, Autobiographical Notes (La Salle & Chicago, Illinois: Open Court, 1979), p.27.

76. See A. Einstein,"Ether and Relativity" in Sidelights on Relativity (London: Methuen, 1922), pp.17-18. See also Einstein, The Meaning of Relativity, pp.54-5; Einstein, "Über den Äther", pp.87-8.

6.7 The Principle of Relativity

Einstein believed that both his special and, in particular, his general theory of relativity had emerged through the elimination of the unacceptable role played by space in earlier physical theories. He had constructed the principle of relativity, in its special and generalised forms as a reaction to this and, through them, believed that he had discovered theories that had led to the establishment of the general relativity of motion. However, as we have seen, Einstein's understanding of the role of space in physical theories, and of the need for its final elimination, changed dramatically in the years following his construction of his general theory of relativity.

Throughout this process Einstein maintained his belief that his theory was still a true theory of general relativity, although it was necessary for him to reinterpret exactly what this meant. In this section I examine the changes in Einstein's understanding of this feature of his theory by surveying the development of the principle which encapsulated it, the principle of relativity.

Einstein's first critique of the role of space in physical theory had come with his 1905 special theory of relativity. There he laid down the requirement that a state of absolute rest was no longer to appear in the laws of physics. This requirement found expression in the principle of relativity. In his 1905 paper, he dealt with the question in the introduction, immediately after his discussion of the magnet-conductor thought experiment. He wrote:

> "Examples of this sort, together with the unsuccessful attempts to discover any motion of the earth relatively to the 'light medium', suggest that the phenomena of

> of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest. They suggest rather that, as has already been shown to the first order of small quantities, the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good. We will raise this conjecture (the purport of which will hereafter be called the 'Principle of Relativity') to the status of a postulate..." [77]

With the special theory of relativity, Einstein believed that a state of absolute rest had been eliminated from the laws of physics. However, certain preferred states of motion remained as a part of these laws and were represented by the set of inertial frames of reference. As an essential part of his programme of establishing the general relativity of motion, he set out to eliminate these as well. In his 1916 Annalen der Physik exposition of his new theory, he dealt with this question with the thought experiment of the two fluid spheres in relative rotation discussed earlier. He regarded the need to appeal to certain preferred states of motion within Newtonian mechanics in order to explain the presence or absence of inertial forces as quite unacceptable. This led him to the following conclusion, where R_1 and R_2 are inertial and rotating frames of reference:

> "Of all imaginable spaces R_1,R_2,etc., in any kind of motion relatively to one another, there is none which we may look upon as privileged *a priori* without reviving the above mentioned epistemological objection. *The laws of physics must be of such a nature that they apply to systems of reference in any kind of motion*. Along this road we arrive at an extension of the postulate of relativity." [78]

---

77. A. Einstein, "On the Electrodynamics of Moving Bodies" in The Principle of Relativity (New York: Dover, 1952), pp.37-8.

78. Einstein, "The Foundation of...", p.113. Einstein's emphasis.

Thus we can say that this extension of the principle of relativity amounts to the requirement that all frames of reference or that all states of motion are physically equivalent, where this is understood to mean that none appear in the laws of nature in any preferred way.

This extended principle must be carefully distinguished from another principle, which we would now call the principle of general covariance. Einstein introduced this principle a little later in the paper as a part of his discussion of the use of arbitrary space-time coordinate systems. It read:

> "The general laws of nature are to be expressed by equations which hold good for all systems of co-ordinates, that is, are covariant with respect to any substitutions whatever (generally covariant)." [79]

This principle bears a strong similarity to the generalised principle of relativity. However there is an important difference. This latter principle requires the physical equivalence of all frames of reference as far as the laws of nature are concerned. The principle of general covariance, however, makes the weaker requirement that all coordinate systems be equivalent as far as the writing of the laws of nature are concerned and, more specifically, that all laws of nature be capable of being written in a form that is the same in all coordinate systems.

The failure to distinguish these principles and their consequences adequately has been the source of some confusion. A generalisation of the principle of relativity does entail an extension of the

---

79. Ibid., p.117.

requirement of Lorentz covariance of the special theory. For if the state of motion of one's reference frame is to no longer enter into the laws of nature as seen in that frame, then these laws must take on the same form when determined by an observer in any state of motion. However, the generalised relativity principle does not entail a complete principle of general covariance, the requirement that the laws of nature retain their form under all coordinate transformations. For, whilst each shift to a new state of motion will involve a transformation of coordinates, each transformation of coordinates does not necessarily involve a shift to a new state of motion.

On the other hand, the establishment of even a complete principle of general covariance does not entail any generalisation of the principle of relativity. For example, the special theory of relativiity can be written in a generally covariant form, yet it does not satisfy any extension of the principle of relativity. Fock in particular has stressed this point and has argued that the general covariance of Einstein's theory does not of itself mean that it is a true theory of general relativity. He has also argued that Einstein believed that his theory was such a theory on the basis of exactly this consideration.[80] We shall see that Einstein's later views on this question were far more sophisticated than this. In any case it is clear that the principles of general relativity and of general covariance should not be confused. The latter principle operates on a "meta" level and the fact that a theory of mechanics satisfies its requirements in no way guarantees that the theory is a true

---

80. Fock, pp.350-2.

theory of general relativity.[81]

Whether Einstein confused the two principles in his 1916 paper is not completely clear. For he continued from the statement of the principle of general covariance given above to conclude: "It is clear that a physical theory which satisfies this postulate will also be suitable for the general postulate of relativity."[82] This shows that he drew a distinction between the two principles and suggests that he recognised that a theory which satisfied the principle of general covariance might also contain a generalised principle of relativity, but need not do so. This is reinforced by the way Einstein continues to justify the principle. This justification is not based on the epistemological considerations used to lead up to the generalised principle of relativity. Rather it is based on the claim that our dealings with processes in spacetime involve only the determination of spacetime coincidences. The choice of coordinate system only enters in the question of the description of these coincidences and should not enter into the laws that govern these processes themselves.

However, there is also some evidence that he did confuse the import of the two principles, for in the same paragraph he noted that "...this requirement of general co-variance ... takes away from space and time the last remnant of physical objectivity...". In any case, in his popular exposition of the theory written in that same year, he not only confused the two principles but identified

81. See E. Zahar, "Einstein, Meyerson and the Role of Mathematics in Physical Discovery", British Journal for the Philosophy of Science, 31 (1980), pp.41-2.

82. Einstein, "The Foundation of ...", p.113.

them. After a discussion of the use of arbitrary Gaussian coordinates in spacetime he returned to the "general principle of relativity" which he had discussed earlier, to give it an "exact formulation".

"We are now in a position to replace the provisional formulation of the general principle of relativity given in Section XVIII by an exact formulation. The form used there, 'All bodies of reference K,K', etc., are equivalent for the description of natural phenomena (formulation of the general laws of nature), whatever may be their state of motion', cannot be maintained, because the use of rigid reference-bodies, in the sense of the method followed in the special theory of relativity, is in general not possible in space-time description. The Gauss co-ordinate system has to take the place of the body of reference. The following statement corresponds to the fundamental idea of the general principle of relativity:
'All Gaussian co-ordinate systems are essentially equivalent for the formulation of the general laws of nature'.

"We can state this general principle of relativity in still another form, which renders it yet more clearly intelligible than it is when in the form of the natural extension of the special principle of relativity. According to the special theory of relativity, the equations which express the general laws of nature pass over into equations of the same form when, by making use of the Lorentz transformation, we replace the space-time variables x,y,z,t, of a (Galilean) reference-body K by the space-time variables x',y',z',t', of a new reference-body K'. According to the general theory of relativity, on the other hand, by application of arbitrary substitutions of the Gauss variables x_1, x_2, x_3, x_4, the equations must pass over into equations of the same form; for every transformation (not only Lorentz transformation) corresponds to the transition of one Gauss co-ordinate system into another." 83

In this passage Einstein explains why this reformulation was made. To begin with, the generalised principle of relativity was seen as a direct extension of the principle of relativity of the special theory. This original principle required the equivalence of all

83. Einstein, Relativity. The Special and the General Theory, pp.97-8. Einstein's emphasis.

inertial frames of reference with regard to the laws of nature and these frames of reference were visualised as rigid bodies extended in space. Now the concept of such a rigid body is, strictly speaking, already inadmissible in the special theory and, as Einstein noted above, unusuable in the most general cases involved in a theory of general relativity. Thus a more rigorous formulation of the principle was needed. The original principle of relativity had translated into a requirement of Lorentz covariance, when it came to the actual manipulation of the special theory of relativity. So, correspondingly, Einstein took the questionable expedient of reformulating the generalised principle of relativity as a principle of general covariance.

This was the form which Einstein gave to the generalised principle of relativity in his important 1918 Annalen der Physik paper. There he presented the foundations of the theory in the form of three principles: the principle of relativity, the principle of equivalence and Mach's principle. The first read:

> "Principle of Relativity: The laws of nature are only assertions about timespace coincidences therefore they find their only natural expression in generally covariant equations." [84]

In the last section we saw that Einstein appended a footnote to the third principle, Mach's principle, in which he noted that he had hitherto not distinguished this principle from the principle of relativity, which he believed had caused confusion. This statement is a little puzzling for, as we have seen, Einstein's earlier work

84. "Relativitätsprinzip: Die Naturgesetze sind nur Aussagen über zeiträumliche Koinzidenzen; sie finden deshalb ihren einzig natürlichen Ausdruck in allgemein kovarianten Gleichungen." Einstein, "Prinzipielles zur...", p.241.

did contain statements of a generalised relativity principle virtually identical to the one given here and which would not be confused with his requirement of the relativity of inertia.

Perhaps Einstein's comment can be understood best by looking at his concept of general relativity in a wider context. Up to 1916 Einstein approached the problem of general relativity within a corpuscular ontology, in which the only causally efficacious entities of the universe were discrete bodies. In such a context, the establishment of the generalised principle of relativity amounted to the establishment of the general relativity of motion of these bodies. As we have seen, in this context, such a result in turn entails the relativity of inertia. Thus, up to 1916, the two requirements of the generalised principle of relativity and the relativity of inertia would have come hand in hand inevitably and it would have been natural to regard them as belonging together.

However, by 1918, Einstein no longer regarded the relativity of inertia with its action at a distance connotations as a requirement belonging to the theory. But he did try to salvage what he regarded to be the essential contents of the requirement in the new principle which he called "Mach's principle". This new principle, in the field ontology in which Einstein was coming to work, was no longer entailed by the generalised principle of relativity. Thus there arose the need to recognise its independent status and to restate it as a separate principle in the axiomatic foundations of the theory.

This development shows the inappropriateness of Einstein's attempt to transplant his original idea of the general relativity of motion and the relativity of inertia to the new conceptual framework. What had first appeared as essentially a single requirement was becoming fragmented and now had to be introduced as two distinct requirements into the axiomatic foundations of the theory.

Thus, in 1918, Einstein divided the work of guaranteeing that his theory was a true theory of general relativity between these two principles, the principle of relativity and Mach's principle. I have already argued that the principle of relativity cannot contribute to this process as long as it is stated in the form given by Einstein in 1918, that is, as the principle of general covariance. At this time, Einstein was under pressure on this question and it was Kretschmann who pressed exactly this point. As a part of a lengthy critique of Einstein's theory he claimed that the principle of general covariance was physically empty, since, he argued, any theory could be rewritten in a generally covariant form purely by mathematical transformation without altering its physical content. [85]

In his 1918 paper, Einstein replied in detail only to this particular point of Kretschmann's:

> "I consider Hr. Kretschmann's argument to be correct, but the innovation proposed by him not to be commendable. That is to say, if it is also correct that one must be able to bring each empirical law into a generally covariant form, then principle (a) [the principle of relativity] still possesses a meaningful heuristic force, which has proven itself brilliantly indeed already on the problem of gravitation and rests on the following. Of two theoretical systems compatible with experience, the one will be preferred which, from the standpoint of the absolute differential calculus, is the simpler and more transparent. If one only brings Newton's gravitational mechanics into the form of absolute covariant equations (four

85. E. Kretschmann, "Über den physikalischen Sinn der Relativitätspostulate, A. Einsteins neue und seine ursprüngliche Relativätstheorie", Annalen der Physik, 53 (1917), pp.575-614.

dimensional), one will be convinced surely that this theory excludes principle (a), certainly not theretically but practically!" [86]

This admission served to further weaken Einstein's theory as a theory of general relativity. For whether the satisfaction of the requirement of the generalised principle of relativity, as stated in the 1918 paper, could be used to support the contention that the theory was a true theory of general relativity was no longer clear. Einstein even seemed to admit the possibility of a generally covariant rewriting of Newtonian gravitation theory. He seemed to argue that it was not the fact that a theory could be written in a generally covariant fashion which guaranteed that the theory was a true theory of general relativity. Rather, it was that the theory took on an especially simple form when rewritten in this way. Thus the theory selection criterion involved in the requirement of general covariance came to be based ultimately on the demand for simplicity, a demand whose meaning and status is far from clear. [87]

---

86. "Ich halte Hrn. Kretschmanns Argument für richtig, die von ihm vorgeschlagene Neuerung jedoch nicht für emphehlenswert. Wenn es nämlich auch richtig ist, dass man jedes empirisches Gesetz in allgemein kovariante Form muss bringen können, so besitzt das Prinzip(a) doch eine bedeutende heuristische Kraft, die sich am Gravitationsproblem ja schon glänzend bewährt hat und auf folgendem beruht. Von zwei mit der Erfahrung vereinbarten theoretischen Systemen wird dasjenige zu bevorzugen sein, welches vom Standpunkte des absoluten Diffentialkalküls das einfachere und durchsichtigere ist. Man bringe einmal die Newtonsche Gravitationsmechanik in die Form von absolut kovarianten Gleichungen (vierdimensional) und man wird sicherlich überzeugt sein, dass das Prinzip (a) diese Theorie zwar nicht theoretisch, aber praktisch ausschliesst!" Einstein, "Prinzipielles zur...", p.242.

87. Einstein still presents essentially the same view some thirty years later in his Autobiographical Notes, op.cit., p.65.

But, in 1918, the third principle in the foundations of the theory Mach's principle, could still be seen to make some contribution to the theory as a theory of general relativity. For, as I have argued earlier, the 1916 version of the theory could not be regarded as a complete theory of general relativity in itself, because preferred states of motion could be introduced as boundary conditions to solutions to the field equations. Mach's principle demanded the exclusion of such boundary conditions and Einstein's implementation of the principle involved the assumption that the universe is spatially closed.

However, as we have seen in the last section, Mach's principle and the ideas surrounding it met with a number of difficulties and Einstein was ultimately led to drop them, once again compromising the idea of his theory as a ture theory of general relativity. Also, at the same time, Einstein came to recognise that his theory by no means removed the last trace of physical objectivity from space, as his original understanding of the general relativity of motion had required. Rather it promoted the antithetical idea that space was capable of becoming the sole medium of physical reality.

Indeed the following-through of this field approach to his theory led Einstein to conclude that the concept of a particle and its motion could play no fundamental role in the theory. Thus he wrote in 1950:

> "Since the general theory of relativity implies the representation of physical reality by a <u>continuous</u> field, the concept of particles or material points cannot play a fundamental part, nor can the concept of motion. The particle can only appear as a limited region in space in which the field strength

or the energy density are particularly high."[88]

In the face of such a conclusion, it is clearly quite impossible to retain the notion that the general relativity of the motion of bodies is a fundamental concept of the theory, for the concept of a particle and its motion no longer even appear as basic terms in the theory.

Thus over the years following 1916, the ideas which supported Einstein's interpretation of his theory as a theory of general relativity were to almost completely break down. But Einstein never lost his conviction in the correctness of the essential ideas of the critique which supported his original interpretation. He began to develop a new account of the essential achievements of his theory and summarised it as a part of a 1927 tribute to Newton:

> "The general theory of relativity formed the last step in the development of the programe of the field-theory. Quantitatively it modified Newton's theory only slightly, but for that all the more profoundly qualitatively. Inertia, gravitation and the metrical behaviour of bodies and clocks were reduced to a single field quality; this field itself was again postulated as dependent on bodies (generalization of Newton's law of gravity or rather the field law corresponding to it, as formulated by Poisson). Space and time were thereby divested not of their reality but of their causal absoluteness - i.e.affecting but not affected - which Newton had been compelled to ascribe to them in order to formulate the laws then known." [89]

In his 1921 Princeton lectures Einstein related the key idea of this account, the elimination of the causal absoluteness of space, to his original critique of the origin of inertial forces:

---

88. A. Einstein, "On the Generalized Theory of Gravitation", in Ideas and Opinions (London: Souvenir, 1973), p.348. Einstein's emphasis.

89. A. Einstein, "The Mechanics of Newton and their Influence on the Development of Theoretical Physics" in Ideas and Opinions (London: Souvenir, 1973), p.260.

"...it is contrary to the mode of thinking in science to conceive of a thing (the space-time continuum) which acts itself, but which cannot be acted upon. This is the reason why E. Mach was led to make the attempt to eliminate space as an active cause in the system of mechanics. According to him, a material particle does not move in unaccelerated motion relatively to space, but relatively to the centre of all the other masses in the universe; in this way the series of causes of mechanical phenomena was closed, in contrast to the mechanics of Newton and Galileo. In order to develop this idea within the limits of the modern theory of action through a medium, the properties of the space-time continuum which determine inertia must be regarded as field properties of space, analogous to the electromagnetic field. The concepts of classical mechanics afford no way of expressing this. For this reason Mach's attempt at a solution failed for the time being."[90]

Even in his last accounts of the significance of the general theory of relativity, Einstein retained this reinterpretation of his earlier views. The theory was no longer seen to eliminate space from physics. Rather it was seen to eliminate the objectionable feature of space, that it acted but was not acted upon, and he now said that it was against this feature of space that his original critique had been directed.[91]

Einstein never made entirely clear what his objection was to this "absolute" character of space. The fact that such a space would act without being acted upon certainly sets it aside from other entities known in physics. But per se this cannot amount to a fatal objection. A more complete explanation can be constructed by recalling that Einstein's original critique of space, which he related to his later critique, was based on the principle of sufficient reason.

---

90. Einstein, The Meaning of Relativity, pp.54-5.

91. See Einstein, Autobiographical Notes, pp.25-7; Einstein, "On the Generalized Theory...", p.348; Speziali, pp.267-8.

Existing theories, he felt, could not provide sufficient reason for certain frames of reference being inertial and others not. In this form, the objection is closely related to a principle of physical determinism. [92] It amounts to the rejection of the earlier theories because they do not enable an account to be given of the causes which determine the frames of reference which are to be inertial. Thus, in those cases in which space acts through its inertial frames, these theories do not provide the opportunity for an account to be given of the causes which completely determine the effects produced, in contradiction to the requirements of a principle of physical determinism. If, however, space could be seen to be acted on as well as acting, then there would be at least the possibility of uncovering the causes which determined the frames that were to be inertial.

To a certain extent, the general theory of relativity is no longer prone to such objections. On the one hand space can be said to act on matter by the medium of inertial frames of reference. On the other hand, matter acts back on space through the field equations and, in this sense, the disposition of inertial frames of reference is influenced by the matter of the universe. Presumably it was this type of behaviour which Einstein found noteworthy in the earlier system which he attributed to Mach, in the above quote, and in which "...the series of causes of mechanical phenomena was closed...".

However, in the general theory of relativity this series of causes is not entirely closed. For their is usually a need for the introduction of boundary conditions in solutions of the field equations,

92. See Dorling, op.cit.

whose nature directly contradicts the requirements of Einstein's critique. We have seen how Einstein sought to avoid this problem with Mach's principle, which required that the metric field be completely determined by its sources, and with the introduction of a closed universe. This attempt was not successful, but the overall ideas contained within it were carried forward into Einstein's search for a unified field theory.

In this search Einstein sought to break down the dichotomy between the metric field and its sources. They were to be fused into a single entity which would comprise the metric field and all forms of matter. In such a system, the requirement of closure of the series of causes of mechanical phenomena amounted to the requirement that this unified field be a closed and self sustaining entity. So, in 1953, Einstein could write:

> "The victory over the concept of absolute space or over that of the inertial system became possible only because the concept of the material object was gradually replaced as the fundamental concept of physics by that of the field. Under the influence of the ideas of Faraday and Maxwell the notion developed that the whole of physical reality could perhaps be represented as a field whose components depend on four space-time parameters. If the laws of this field are in general covariant, that is, are not dependent on a particular choice of co-ordinate system, then the introduction of an independent (absolute) space is no longer necessary. That which constitutes the spatial character of reality is then simply the four-dimensionality of the field. There is then no 'empty' space, that is, there is no space without a field." [93]

---

93. A. Einstein, Foreword to M. Jammer, Concepts of Space, op.cit., p.xv. See also J. Callaway, "Mach's Principle and Unified Field Theory", Physical Review, 96 (1954), pp.778-80.

This outlines Einstein's final understanding of his theory and its significance for the concept of space. It is far removed from Einstein's original vision of the general relativity of motion, the relativity of inertia and the final elimination of space. But we can see how each of these concepts contributed to his final view: the general relativity of motion yielded the concept of general covariance, the relativity of inertia yielded the idea that the unified field should be an entirely self sufficient entity and the elimination of space as an undetermined causal entity yielded to what amounted to the elevation of space to a central, interactive role in the unified field.

6.8 Conclusion

In 1912 and 1913 Einstein made the crucial step from his scalar theory of gravitation to the basic framework of the general theory of relativity. This step was mediated by the expectation that this new theory would yield an extension of the relativity of motion from the relativity of inertial motion, established by his 1905 special theory of relativity, to the relativity of all motion. This would complete the elimination of space as a causally active agent in physical theory. This process had begun with the special theory of relativity, which Einstein believed had banished a state of absolute rest from the laws of nature. He hoped that his new theory would eliminate any preferred state of motion from the laws of nature and, in particular, the prior status of the set of inertial frames of reference used to account for the origin of inertial forces.

However, in 1913, Einstein was not certain that his new theory had established the general relativity of motion, for he had been unable to discover generally covariant field equations, although the remainder of the theory was generally covariant. Nevertheless, he was confident that he was close to a true general theory of relativity. For he believed that the elimination of inertial frames of reference as the irreducible source of inertial forces required that their origin now be seen to lie in an inductive interaction between the accelerated body experiencing the forces and all the remaining bodies of the universe. He called this the hypothesis of the relativity of inertia and he found that he could recover certain weak field effects from his theory which were consistent with it. Further, he was able to use the requirement as a criterion for rejecting Nordström's theory of gravitation, the theory which he saw as the strongest competitor to his own theory, without the need to invoke the requirement of the general relativity of motion directly.

Einstein's confidence in the theory as a theory of general relativity grew with his temporary seduction by his notorious proof that generally covariant field equations contradicted the requirements of causality and with the 1915 discovery of generally covariant field equations. With this came a new confidence in his understanding of the foundations of the theory. Einstein saw the theory as following from a defect in earlier theories, which was not just empirical but epistemological in nature, and whose discovery in all its clarity was attributed by him to Mach. This involved a mistaken interpretation of the import of Mach's critique, but, in itself, need not have compromised Einstein's theory. In any case, out of this critique

Einstein drew the requirements of the relativity of inertia and the general relativity of motion, which he presented as essential to the foundations of his theory.

In retrospect, it is by no means clear that the 1915 version of the theory was a true general theory of relativity, for it was possible and usually necessary to introduce boundary conditions to solutions of the field equations which contradicted this requirement. Over the years following, Einstein turned to face this difficulty and, moreover, came to recognise that his new theory promoted concepts which led to contradictions with his original understanding of the foundations of the theory. He had developed the concept of the general relativity of motion and the relativity of inertia within a world view which saw the universe as consisting of discrete bodies in motion in the featureless container of space. In particular, an essential part of his critique was the implicit assumption that these bodies were the only causally efficacious entities in the universe. This last assumption became untenable with Einstein's growing awareness that his theory of relativity promoted the concept of the reality and even the primacy of fields, causal entities which permeate space. Further, as part of this process, he came to see that his theory did not eliminate space from any active role in the laws of nature. Rather, it was elevated to the status of one of the most important causally active fields, the metric field.

First Einstein sought to deal with the boundary condition problem and retain the essential idea of the relativity of inertia in field theoretic terms with the introduction of what he called Mach's principle and the idea of a spatially closed universe. However, this episode met with little success. He also began to weaken the generalised principle of relativity, which had expressed his original idea of the complete elimination of space and the physical equivalence of all states of motion. It became, in effect, the requirement that the formal statements of all physical laws be capable of generally covariant expression, a requirement which, from as early as 1917, has been claimed to have no physical content.

Thus the years following 1916 saw an almost complete breakdown of the concepts underpinning Einstein's original understanding of his theory. Out of this, however, there emerged a new concept to replace them - the unified field - an all-embracing and self-sufficient entity which was to contain the known fields of gravitation and electromagnetism, all known matter and the metric field of spacetime as well. This concept began to eclipse Einstein's original idea of the general relativity of motion. He still retained his belief that his early critique had uncovered a fundamental defect in the then existing theories which was removed by his general theory of relativity. But this was no longer seen to be achieved by the complete elimination of space from the laws of nature. Rather it came through the recognition that space, as a part of the unified field, was an entity that could be acted upon as well as acting itself. Thus, physical theory would no longer be forced to regard space as an undetermined first cause at the beginning of a chain of causal actions. Rather, the chain could now be traced back beyond it to the causes which act on it as well.

But in spite of these later developments, it is clear that Einstein's original idea of the general relativity of motion played a crucial role in the historical development of his general theory of relativity. It functioned as an important heuristic to direct Einstein and give him confidence in his developing theory. Further, it caused him to seek a generally covariant formalism for dealing with gravitation theory. This led him to the work of Riemann, Christoffel, Ricci and Levi-Civitã and thus to the use of generally covariant tensors in a Riemannian spacetime as the basic terms of his gravitation theory.

CHAPTER 7

EINSTEIN'S THEORY AS A THEORY OF GRAVITATION

7. EINSTEIN'S THEORY AS A THEORY OF GRAVITATION

7.1 An Historical Problem

In this chapter I turn to face a problem which has been brewing in the preceding chapters. Einstein's path to his final general theory of relativity was long and tortuous. From his own accounts it would appear that his ultimate success was only made possible by the guidance at crucial intersections of a small number of fundamental and far-reaching insights: the inability of special relativity to contain gravitation, the equivalence of acceleration and gravitation and the concepts of the elimination of space and the general relativity of motion. Again and again in his journey to the final theory we find Einstein returning to and insisting on the fundamental significance of these insights. Yet key aspects of each of these insights are inconsistent with the final theory which Einstein produced.

In 1907 Einstein had rejected the possibility of an acceptable special relativistic gravitation theory. Whilst later developments might well be seen to bear this out, his justification for this at the time was quite incomplete. It was based on an insistence on the independence of the acceleration of a falling body from its sideways velocity. This independence, it turns out, is not a part of the final general theory of relativity. But exactly such an independence can be recovered from a second rank tensor gravitation theory still within special relativity.

Following this, Einstein postulated the equivalence of acceleration and gravitation, within which he saw the promise of an extension of the relativity of motion to accelerated motion. In the form of the equivalence of uniform acceleration and homogeneous gravitational fields, it became the core of his new relativistic gravitation theory in the years prior to 1913. He hoped to use it as a bridge which would enable him to pass from the accessible structure of inertial fields to the more obscure structure of gravitational fields. We have seen that Einstein's final general theory of relativity severely limits the prospects of such a device. The "homogeneous gravitational fields" involved would no longer be regarded to be gravitational fields at all. Further, the theory does not admit a thoroughgoing equivalence of acceleration and gravitation of the type invoked by Einstein. Whilst inertial fields may mimic the gross features of gravitational fields, the theory provides an unambiguous and local arbiter to distinguish the two, the curvature of spacetime as measured by the Riemann curvature tensor. Therefore, the theory does not allow gravitation to stand for acceleration, in the way Einstein originally expected, and thus does not admit his original conception of the extension of the relativity of motion to accelerated motion.

Finally, and of greatest importance, there was Einstein's critique of space and his demand for its final elimination as a causally active agent from physical theory. This found expression in the search for a theory which would realise the general relativity of motion. It had underpinned the launching of his 1907 Jahrbuch... gravitation theory and, in 1912 and 1913, had led him to seek a generally covariant gravitation theory and to hypothesise the relativity of inertia.

However, we have seen that Einstein's critique was both ill-born and ill-fated. It arose from a misunderstanding of the import and power of Mach's original critique. Further, the critique and some of the most important concepts derived from it - the general relativity of motion and the relativity of inertia - were dependent on a kind of corpuscular world view, in which the only causally active agents were seen to be discrete particles contained within a passive and featureless container space. Einstein's own developing theory of relativity, with its insistence on a field account of nature, was at the same time completing the overthrow of this world view. We have seen how Einstein found it increasingly difficult to sustain his critique and the concepts surrounding it, as his new theory unfolded. Repeatedly he had to reformulate and reassess them and their position in the theory until he was finally led to concede that his general theory of relativity promoted a conception which was antithetical to his original critique - that space may well be the ultimate medium of all reality.

In the light of such developments, the crucial question which must be asked, is this: How was it possible for Einstein to construct a theory which is hailed as one of the greatest achievements of modern theoretical physics and even of human speculative thought as a whole, but which so fundamentally contradicts the basic insights and heuristics used in its construction? In answer, we might attribute the appearance of the final theory to an accident of history. Or, perhaps, we might reaffirm our faith in the certainty with which Einstein's "intuition" or "genius" would lead him safely through the complexities and contradictions to the final successful theory.

If the history of science is to be anything more than a mere catalogue of accidents and mysteries, then neither of these answers is at all acceptable. For they amount to the renunciation of any hope of understanding the origins of Einstein's theory and banish the question of its origins to the realm of the unanalysable.

In this chapter I propose an alternative. Einstein's work on his general theory of relativity came to be dominated by the hope that with it he would arrive at a theory which would fulfil the demands of Mach's critique of Newtonian mechanics, realise the final elimination of space from physical theory and complete the relativistic revolution precipitated by his special theory of relativity. Beneath this grand vision lay the more mundane task of assimilating gravitation into the new relativistic world view. As we have seen, Einstein recognised that this task would lead to a modification of that world view. But, of greater importance, he believed that completion of this task was essential to the successful construction of a true general theory of relativity.

Einstein was driven forward in his search for his final theory by such requirements as the equivalence of gravitation and acceleration and the elimination of space. But each step that he took had to remain within the constraint envelope which arose from his requirement that the final theory contain a relativistically acceptable gravitation theory. In this chapter and the next, I will argue that as long as Einstein remained within that constraint envelope, as I shall continue to call it, and carried through its implications to the full, then he would be drawn inexorably to the general framework of his final theory. This is not to say that the emergence of the theory was inevitable. For the task of recognising exactly what the implications of these constraints were is by no means trivial and it cannot be guaranteed that they would be seen. However, this does make intelligible how it would be possible for Einstein to develop his general theory of relativity in apparent contradiction to the insights and heuristics which gave birth to it.

Thus, I am seeking to characterise Einstein's path to his general theory of relativity in the following way. Einstein was driven forward by a group of powerful but organic insights and heuristics at whose heart lay the hope that he could construct a true general theory of relativity. However, Einstein's path was circumscribed by the constraint envelope which arose from his requirement that the final theory contain a relativistically acceptable theory of gravitation. Irrespective of whatever success he may have had in his first goal, it was his success in the second which ensured that he would arrive at his final theory or one much like it.

I now turn to the task of mapping out the constraint envelope within which Einstein worked and the group of heuristics which drove him forward. We shall see that it was the uncompromised stability of this constraint envelope which ensured the integrity and continuity of Einstein's programme. But it was the driving heuristics which suggested novel concepts and results of great significance well before we might otherwise have expected them to appear.

7.2 Einstein's Programme

In this section I seek to map out the envelope of constraints, which arose from the requirement that the theory contain an acceptable relativistic theory of gravitation, and the hypotheses of the heuristics which drove Einstein forward in the search for a general theory of relativity.

Group A The Constraint Envelope

A1 The Newtonian Limit In the case of weak gravitational fields and low velocities, the theory should reduce to Newtonian gravitation theory.

A2 The Special Relativistic Limit From the very beginning Einstein believed that space and time would not behave in accord with special relativity in the presence of a gravitational field. However, in the case of an infinitely weak gravitational field, the behaviour of space and time was to return to that of special relativity.

A3 Conservation of Energy and Momentum The laws of conservation of energy and momentum must hold within the theory. In particular the conservation is to encompass total energy and momentum, which includes the energy and momentum of the gravitational field.

A4 Democracy of Energy-Momentum (Equality of gravitational and inertial mass) Matter itself is the active and passive source of the gravitational field. In the early static scalar theory it is measured by the quantity energy and in the final dynamic theory by the four-vector quantity energy-momentum. All forms of matter, as measured by their energy or energy-momentum, are to contribute on an equal level to the source term of the gravitational field, including the energy and momentum of the gravitational field itself. Thus, in this sense, a gravitational field is in part its own source. Since all energy has inertia, this may be reformulated as the requirement of the equality of gravitational and inertial mass. This reformulation is necessarily a loose one since, in the final theory, there is no scalar quantity which corresponds to gravitational mass. Quantity of gravitational source is measured by the four-vector energy-momentum.

This list of constraints has much in common with a list of postulates which Einstein suggested a gravitation theory may seek to satisfy in the introduction to his 1913 Vienna address:

> "In the following we give some general postulates, which can be required of a gravitation theory, but need not all be required:
> 1. fulfilment of the law of conservation of momentum and energy;
> 2. equality of the inertial and gravitational mass of closed systems;
> 3. validity of the theory of relativity (in the narrower sense); i.e. the systems of equations are covariant against linear orthogonal substitutions (generalised Lorentz transformations);
> 4. the observable laws of nature do not depend on the absolute value of the gravitational potential (or, correspondingly, of the gravitational potentials). Physically this means the following: The contents of the relations between observable quantities which one can find in a laboratory will not be changed if I bring the whole laboratory into a region of another gravitational potential (constant in space and time)." [1]

---

1. "Im nachfolgenden geben wir einige allgemeine Postulate an, welche an eine Gravitationstheorie gestellt werden können, aber nicht alle gestellt werden müssen:
   1. Erfüllung der Erhaltungssetze des Impulses und der Energie;
   2. Gleichheit der trägen und der schweren Masse abgeschlossener Systeme;
   3. Gültigkeit der Relativitätstheorie (im engeren Sinne); d.h. die Gleichungssysteme sind kovariant gegenüber linearen orthogonalen Substitutionen (verallgemeinerte Lorentz-Transformationen);
   4. die beobachtbaren Naturgesetze hängen nicht ab vom Absolutwerte des Gravitationspotentials (bzw. der Gravitationspotentiale). Es bedeutet dies physikalisch folgendes: Der Inbegriff der Relationen zwischen beobachtbaren Grössen, welche man in einem Laboratorium finden kann, wird dadurch nicht geändert, dass ich das ganze Laboratorium in ein Gebiet von anderem (räumlich und zeitlich konstanten) Gravitationspotential bringe." A. Einstein "Zum gegenwärtigen Stande des Gravitationsproblems", Physikalische Zeitschrift, 14 (1913), p.1250.

Einstein continued to note that all theoreticians agreed to postulate 1 and that postulate 2 was supported experimentally by Eötvös. Also he noted his disagreement with Abraham's theory of gravitation which did not agree with postulate 3 even in the limiting case of a constant gravitational potential. Finally, Einstein confided that we should have less faith in postulate 4, since it was based only on a belief in the simplicity of the laws of nature.

Group B The Driving Heuristics

B1 The Principle of Equivalence Einstein first hypothesised the equivalence of gravitation and acceleration with the hope that it would provide a bridge that would enable him to pass from inertial to gravitational fields. As we have seen he weakened his idea until it was transformed into the statement that inertia and weight are of the same nature. In this form the principle came to contribute the result to the theory that the trajectory of a free particle in spacetime is independent of its nature.

B2 General Relativity The requirement of the general relativity of motion was first invoked by Einstein as a part of his critique of space. It also involved the hypothesising of the relativity of inertia and then Mach's principle. As we have seen, its final contribution to the theory was in the principle of general covariance and the requirement of "no prior geometry".

There is a clear correspondence between the constraints and hypotheses of Groups A and B and the concepts of the "hard core", "protective belt" and "positive heuristic" delineated by Lakatos.[2] The contents of Group A formed an uncompromisable set of hypotheses within the work which led Einstein to his final 1915 theory and therefore could qualify as the hard core of his research programme. The general ideas of Group B could be said to be the positive heuristic of Einstein's programme, for it was these ideas which stimulated and directed Einstein in his research. Also the hypotheses contained in Group B could be looked upon as the protective belt of hypotheses surrounding the hard core, for, as I will argue in a later section, these hypotheses were altered by Einstein, when he face a contradiction in his developing theory, rather than those of Group A.

However, this correspondence can only be maintained partially. In particular, it is clear from Einstein's own accounts that he regarded the hypotheses of Group B as central to his research programme and not as auxilliary hypotheses constructed to protect the contents of Group A. Rather than look upon the contents of Group A as the essential core of his programme, he would have regarded them as a set of subsidiary hypotheses, albeit uncompromisable, but introduced principally to enable his new theory to mesh in with the existing structure of theoretical physics.

7.3 The Sufficiency of Einstein's Programme

I have claimed that, irrespective of his other considerations, we could be sure that Einstein would be directed towards his final general theory of relativity, or one much like it, simply by virtue of the

---

2. I. Lakatos, "Falsification and the Methodology of Scientific Research Programmes" in I. Lakatos and A. Musgrave, eds., Criticism and the Growth of Knowledge (Cambridge University Press, 1977), pp.91-196.

fact that he was seeking a relativistically acceptable gravitation theory. In order establish this it is necessary to first establish a result in the logical structure of Einstein's theory. That is, that when one seeks a relativistically acceptable gravitation theory then one is indeed drawn towards Einstein's theory or one much like it.

In Chapter 3, the first steps which one might take in the relativisation of Newtonian gravitation theory were sketched out. We saw that there were three routes which could be followed, a scalar, vector or second rank tensor theory, depending on the choice of source term for the field and its transformation properties. The vector theory is formally analogous to the theory of electrodynamics. Nowadays, quite apart from any empirical questions, such a vector theory of gravitation is rejected simply on the grounds that it would require normal masses of like gravitational "charge" to repel, as is the case in electrodynamics, rather than attract one another.[3] However, this defect can be corrected by a mere change of sign. A more complete rejection of such a vector theory involves a demonstration that the theory is unacceptable energetically. This was shown as early as 1912 by Abraham. He argued that a state of equilibrium could never be maintained for a system of particles within the theory, for the slightest perturbation would be magnified and not damped by the inductive forces produced and set the whole system in motion. This was related to the negativity of the field energy density, which implied that energy was transported in the opposite direction to that of the propagation of gravitational waves. This negativity had led Maxwell to

---

3. See S.N. Gupta, "Einstein's and Other Theories of Gravitation", Reviews of Modern Physics, 29 (1957), p.334; W.E. Thirring, "An Alternative Approach to the Theory of Gravitation", Annals of Physics, 16 (1961), p.99.

give up a similar gravitation theory some years before.[4]

This leaves scalar and second rank tensor theories. I argued in Chapter 3 that the fact the mass-energy is the source of a gravitational field directs us towards second rank tensor theories and, to begin with, a field equation of the form

$$\Box^2 \varphi^{\mu\nu} = -4\pi G\, T^{\mu\nu} \tag{3.3}$$

where $\varphi^{\mu\nu}$ are the gravitational potentials and $T^{\mu\nu}$ the stress-energy tensor for normal matter. In this form, the equation is incomplete, for the stress-energy tensor, which comprises the source term, only accounts for non-gravitational energy-momentum, whereas we expect the energy-momentum of the gravitational field itself to contribute to the source term as well. The stress-energy tensor of the gravitational field can be calculated from this field equation and added into the source term. This modification of the field equation in turn leads to a modification of the gravitational field stress-energy tensor which then must be added in turn into the source term. It has been shown that if the cyclical procedure begun here is carried through until the final equations converge, then the gravitation theory which results is formally identical with Einstein's general theory of relativity.[5]

However, with this we have not yet completed the recovery of Einstein's theory, for spacetime is still regarded as flat even though

---

4. See M. Abraham, "Neuere Gravitationstheorien", Jahrbuch der Radioaktivität und Elektronik, 11 (1914), pp.475-6; M. von Laue, "Die Nordströmsche Gravitationstheorie", Jahrbuch der Radioaktivität und Elektronik, 14 (1917), pp.270-1.

5. See S.N. Gupta, "Gravitation and Electromagnetism", Physical Review, 96 (1954), pp.1683-5; J. Weber, "Gravitation and Light" in H.Y. Chiu and W.F. Hoffmann, eds., Gravitation and Relativity (New York: Benjamin, 1964), pp.235-7; C.W. Misner, K.S. Thorne, J.A. Wheeler, Gravitation (San Francisco: Freeman, 1970), pp.177-8, p.436.

its structure is no longer directly accessible to measurement by rods and clocks. On the grounds of parsimony, we reject this inaccessible flat spacetime geometry, accept the Riemannian geometry of spacetime, which is given by rods and clocks, and thus acknowledge that gravitation is nothing other than the curvature of spacetime.

Havas has stressed that the danger in this particular approach lies in the serious difficulties which still surround the concept of gravitational field energy-momentum. He reaffirms, however, that Einstein's theory can be regarded as a natural outcome of the attempts to incorporate gravitation into the framework of relativity theory.[6]

A similar analysis can be carried out in the case of scalar gravitation theory. The starting point of such an analysis has been mapped out in Chapter 3. We shall see later that following on from this with certain consistency requirements, we arrive at a theory which once again can be given a generally covariant formulation and in which gravitation can be seen to correspond to the curvature of spacetime. The fact that such a theory can be recovered reinforces the idea that the incorporation of gravitation into relativity theory automatically leads to the curvature of spacetime. What makes it especially significant is that the recovery of this theory was carried out by Nordström and Einstein at the same time as Einstein was developing his general theory of relativity. This will be examined in greater detail in the following chapter.

In the earlier days of relativity theory, this scalar theory was regarded to be somewhat unnatural, for there was no completely

---

6. P. Havas, "Foundation Problems in General Relativity" in M. Bunge, ed., Delaware Seminar in the Foundations of Physics (Berlin: Springer, 1967), pp.124-48.

generally covariant way of writing the theory known.[7] Nowadays such a formulation is available. However, the scalar theory still does not satisfy the requirement of "no prior geometry", unlike Einstein's theory. Nevertheless it has been argued that formally there is little to choose between the two theories and that any such choice can only be made on empirical grounds. As is well known, experience decides in favour of Einstein's theory, in this case.[8]

We can now return to the constraints and hypotheses delineated earlier and which I asserted governed the development of Einstein's research programme. The arguments referred to in this section, which lead uniquely to the general theory of relativity and the associated generally covariant formulation of Nordström's theory, made use only of requirements drawn from the constraint enevelope of Group A. In particular explicit use was made of the requirements of the Newtonian and special relativistic limits and the democracy of energy-momentum. But no recourse was made to any of the hypotheses of Group B, the principle of equivalence and the general relativity of motion.

Thus we can say that the contents of the constraint envelope are sufficient in a logical sense to generate Einstein's final theory, or one much like it - in other words Nordström's theory. In this sense we can say that the only additional input needed to ensure the unique generation of Einstein's final theory is some reason for choosing it

---

7. See W. Pauli, Theory of Relativity (Oxford: Pergamon, 1958), p.144.

8. See Misner et al., pp.429-31; J.L. Pietonpol and D. Speiser, "Remarks on the Foundations of General Relativity", Helvetica Physica Acta, 48 (1975), pp.153-61.

over Nordström's.

However, in the actual historical process of discovery, Einstein was able to draw many of the crucial features of the final theory from the hypotheses of Group B. This process has already been examined at length in the preceding chapters. For example, we have seen how these hypotheses led Einstein to recognise that space and time does not behave in accord with special relativity in the presence of a gravitational field and that this could be handled by a generally covariant theory which dealt with the curvature of a metrical spacetime. Further, we have seen how these hypotheses acted as a driving force for Einstein as he worked towards his final theory. It provided a framework which suggested new concepts and techniques to assist him in the development of his theory and against which he could assess the success of the various stages of its development.

Finally, it was this group of hypotheses which led Einstein to a second rank tensor theory, rather than a scalar theory. For, as we have seen, it was out of his attempts to develop a generally covariant theory that he was led to represent the gravitational field, as well as the metrical properties of spacetime, by the second rank metric tensor.

7.4 The Shaping of Einstein's Theory

In this section I seek to reinforce the idea of the constraint envelope of Group A influencing the direction of Einstein's programme and controlling the form of the theory which finally emerged. To begin with I need hardly stress Einstein's unwaivering committment to the constraints of Group A. The strength of Einstein's committment tended to be hidden behind his more prominent discussion of such questions as the equivalence of acceleration and gravitation and the general relativity of motion. However, it did emerge on a number of occasions.

For example, in the questions following Einstein's 1913 Vienna address, Reissner challenged Einstein on what I have called the democracy of energy-momentum in the new theory, in particular with regard to the status of gravitational field energy.[9] Einstein misunderstood Reissner's question and therefore returned to the question with a retrospective reply in a later issue of the Physikalische Zeitschrift.[10] We can gauge the importance which Einstein attached to the matter from the length of his reply. He devoted three pages to it in the Physikalische Zeitschrift. His entire Vienna address, which gave an account of both his own theory and Nordström's gravitation theory, took only thirteen pages in the same journal.

Einstein began his reply be reasserting the requirement of the conservation of energy-momentum: "First I recall that it must be unconditionally demanded that matter and energy together satisfy the laws of conservation of momentum and energy."[11] Then he turned to the democracy of energy-momentum. He showed that in his theory the matter of the gravitational field, as represented by the gravitational field stress-energy tensor, behaved in exactly the same way as non-gravitational matter. It was acted on by gravitational fields as well as

---

9. See Einstein, "Zum gegenwärtigen Stande...", p.1265.

10. A. Einstein, "Nachträgliche Antwort an Herrn Reissner", Physikalische Zeitshrift, 15 (1914), pp.108-110.

11. "Zunächst erinnere ich daran, dass unbedingt gefordert werden muss, dass Materie und Energie zusammen den Erhaltungsätzen des Impulses und der Energie genügen." Ibid., p.108.

being field producing itself; it had inertia and entered into the conservation of energy-momentum exactly as did normal matter.

Another case in which Einstein had occasion to stress the importance of the constraints of Group A, is in the vitriolic 1912 dispute between himself and Abraham. Abraham maintained that Einstein's 1912 gravitation theory amounted to a rejection of the special theory of relativity. Einstein, however, staunchly defended his earlier theory as a boundary case of his gravitation theory, as is required by the constraint enevelope. He also felt compelled to reject Abraham's own gravitation theory on the grounds that it did not contain the special theory of relativity even as a limiting case. [12]

Such examples can be multiplied. However, in the following I turn to the more detailed study of two cases of special significance. In these cases Einstein found himself driven forward by the hypotheses of Group B into a collision with the constraint enevelope of Group A. We shall see that at these points of collision Einstein chose to be deflected back into the constraint enevelope by allowing modifications to the contents of Group B, in order to avoid any compromise of the contents of Group A.

This choice explains how Einstein's treatment of the considerations of Group B, the principle of equivalence and the question of general relativity, could alter so dramatically in the course of the development and later elaboration of the final theory, without destroying the integrity of his research programme. The coherence and identity of

---

12. See Einstein, "Zum gegenwärtigen Stande...", pp.1250-1, as well as the references cited in Section 3.4.

this programme was maintained by the unchanging and uncompromisable constraint envelope of Group A. This in turn provides a stable base from which we can review and understand the development of Einstein's theory.

The two cases which follow are important in themselves, for they represent turning points in Einstein's programme. However, I would also like to stress their wider significance. For they demonstrate the power and primacy of the constraints of Group A and their ability to modify the contents of Group B and redirect the course of Einstein's programme.

7.4.1 General Covariance and the "Entwurf..." Field Equations

In 1913, in his "Entwurf..." paper, Einstein found himself driven by the considerations of Group B into what appeared to be a headlong collision with the requirements of the constraint envelope of Group A. It seemed to follow from the requirement of the general relativity of motion that his new theory should contain only generally covariant laws. But, to his dismay, neither he nor Grossmann could find generally covariant field equations for the gravitational field.

In his part of the paper Grossmann noted that differential geometry naturally suggested what we now call the Ricci tensor, the second rank contraction of the fourth rank Riemann curvature tensor, as the basis of a generally covariant field equation. But this possibility was rejected with the brief remark:

> "However it turns out that this tensor does not reduce to the expression $\Delta\varphi$ in the special case of an infinitely weak static gravitational field." [13]

In other words, this possiblity was rejected because it appeared not to reduce to the correct Newtonian limit, the Laplacian of the scalar gravitational potential, $\Delta\varphi$. With this Einstein and Grossmann gave up hope of discovering generally covariant field equations. However, as we saw in the last chapter, it was not until later that Einstein came to the conclusion that such field equations would be inadmissible and presented his notorious "hole" argument as final proof.

Another constraint from Group A also played a role in the development of Einstein's belief that no generally covariant field equations were possible. He recognised that the conservation of energy and momentum must now encompass gravitational energy-momentum as well and thus should take on the following form

$$\left(\sqrt{-g}\,\left(T_{\mu}^{\nu} + t_{\mu}^{\nu}\right)\right)_{,\nu} = 0 \qquad (7.1a)$$

where T_{μ}^{ν} is the stress-energy tensor for non-gravitational matter, t_{μ}^{ν} the stress-energy tensor for the gravitational field and g the determinant of the metric tensor. Einstein recognised that the t_{μ}^{ν} derived from his "Entwurf..." field equations was not a generally covariant tensor and this in turn limited the coordinate systems in which his theory could hold. [14] Einstein soon developed this idea

---

13. "Allein es zeigt sich, dass sich dieser Tensor im Spezialfall des unendlich schwachen statischen Schwerefeldes nicht auf den Ausdruck $\Delta\varphi$ reduziert." A. Einstein and M. Grossmann, "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation", Zeitschrift für Mathematik und Physik, 62 (1913), p.257. Grossmann's emphasis.

14. See Einstein, "Zum gegenwärtigen Stande...", p.1264.

further. He concluded that it was not possible to have a generally covariant conservation law of the above form and that this automatically led to a specialisation of coordinate systems, which would in turn be reflected in the limited covariance of the field equations. [15]

Even though Einstein and Grossmann were mistaken on both these questions, they show quite clearly the power of the constraint envelope over the direction of Einstein's research programme and its primacy over the contents of Group B. For rather than break out of the constraint enevelope, Einstein was prepared to suspend and ultimately give up the requirement of general covariance, a requirement which was later re-established and repeatedly reaffirmed as one of the fundamental concepts of the final theory.

But the episode did not end here. For Einstein found that the requirement that his gravitation theory reduce to a Newtonian limit and also contain a combined law of conservation of gravitational and non-gravitational energy-momentum led him to a unique set of field equations, the field equations of the "Entwurf..." paper. Einstein presented a detailed derivation of these equations only in his "Entwurf..." paper. [16] But he repeated his belief in a number of other places that they were fully determined by the above constraints in conjunction with some natural simplicity requirements. [17]

---

15. A. Einstein, "Prinzipielles zur verallgemeinerten Relativitätstheorie", Physikalische Zeitschrift, 15 (1914), p.178. See also A. Einstein, "Physikalische Grundlagen einer Gravitationstheorie", Naturforschende Gesellschaft Vierteljahrsschrift (Zürich), 58 (1913), pp.288-9.

16. Einstein, "Entwurf...", pp.233-9.

17. See Einstein, "Physikalische Grundlagen...", pp.288-9; Einstein, "Zum gegenwärtigen Stande...", p.1258.

In the "Entwurf..." paper Einstein went to great pains to explain the ingenious method which he had used to construct his field equations. To make it quite clear, he first used his method to show how one could derive the field equation of electrostatics and only then turned to the gravitational field equations. In the following I revert to the notation of the "Entwurf..." paper.*

Einstein began by assuming that the equations sought were of the form

$$\Gamma_{\mu\nu} = \varkappa \Theta_{\mu\nu} \tag{7.2a}$$

where $\Theta_{\mu\nu}$ is the stress-energy tensor for non-gravitational matter, $\varkappa$ is a constant and $\Gamma_{\mu\nu}$ is a second rank tensor composed of derivatives of the metric tensor no higher than the second. This equation was a natural generalisation of Poisson's equation in Newtonian theory. Einstein noted that it had proved impossible to find a generally covariant set of field equations under these conditions and raised the possibility that such equations might be discovered if higher derivatives of the metric tensor were considered. This question was dismissed with a brief comment:

> "The attempt of discussion of such possibilities would however lead us astray with the current status of our knowledge of the physical properties of gravitational fields." [18]

---

* Greek indices take values from 1 to 4, all being written "down". Covariant tensors are indicated by Latin letters and their corresponding contravariant forms by the corresponding Greek letter. Gothic letters represent mixed tensor densities. The Einstein summation convention is not used.

18. "Der Versuch einer Diskussion derartiger Möglichkeiten wäre aber bei dem gegenwärtigen Stande unserer Kenntnis der physikalischen Eigenschaften des Gravitationsfeldes verfrüht." Einstein, "Entwurf...", p.234.

Einstein had already derived the law of conservation of energy-momentum as the vanishing of the covariant divergence of the (non-gravitational) stress-energy tensor. This he wrote in the expanded form

$$\sum_{\mu\nu} \frac{\partial}{\partial x_\nu}\left(\sqrt{-g}\; g_{\sigma\mu}\Theta_{\mu\nu}\right) - \frac{1}{2}\sum_{\mu\nu}\sqrt{-g}\,\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\,\Theta_{\mu\nu} = 0 \qquad (7.1b)$$

where $g_{\mu\nu}$ is the metric tensor and g its determinant. Einstein noted that the second term of this equation represented the transfer of energy-momentum from the gravitational field to matter. Thus it could be expected to appear in a law

$$\sum_{\mu\nu} \frac{\partial}{\partial x_\nu}\left(\sqrt{-g}\; g_{\sigma\mu}\,\theta_{\mu\nu}\right) = \frac{1}{2}\sum_{\mu\nu}\sqrt{-g}\,\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\,\Theta_{\mu\nu} \qquad (7.1c)$$

where $\theta_{\mu\nu}$ represents the stress-energy tensor for the gravitational field. Therefore the left hand side of this equation must be composed only of the metric tensor and its derivatives. Einstein recognised this last conclusion at the beginning, but did not explicitly associate this side of the equation with the gravitational field stress-energy tensor until later in the paper. Equation (7.1c) is of course merely a restatement of the law of conservation of energy-momentum and can be constructed directly from the two other forms of the law given in equations (7.1a) and (7.1b). Now, Einstein continued, if $\Theta_{\mu\nu}$ in equation (7.1c) is replaced by $\frac{1}{\kappa}\Gamma_{\mu\nu}$ in accord with the field equations we have

$$\sum_{\mu\nu} \frac{\partial}{\partial x_\nu}\left(\sqrt{-g}\; g_{\sigma\mu}\,\theta_{\mu\nu}\right) = \frac{1}{2}\sum_{\mu\nu}\sqrt{-g}\,\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\,\Gamma_{\mu\nu} \qquad (7.3a)$$

which must be a set of identities containing only the metric tensor and its derivatives.

Then came the crucial step. Einstein argued that if one did not know the field equations, all one had to do was discover a set of identities in the metric tensor and its derivatives of the appropriate form and from this one could then recover a set of field equations. In his mathematical part, Grossmann had constructed suitable identities and Einstein presented them as "uniquely determined", thus implying that the resulting field equations were also uniquely determined.[19] The identities were:

$$\sum_{\alpha\beta\tau\rho} \frac{\partial}{\partial x_\alpha}\left(\sqrt{-g}\, \gamma_{\alpha\beta} \frac{\partial \gamma_{\tau\rho}}{\partial x_\beta} \frac{\partial g_{\tau\rho}}{\partial x_\sigma}\right) - \frac{1}{2}\sum_{\alpha\beta\tau\rho} \frac{\partial}{\partial x_\sigma}\left(\sqrt{-g}\, \gamma_{\alpha\beta} \frac{\partial \gamma_{\tau\rho}}{\partial x_\alpha} \frac{\partial g_{\tau\rho}}{\partial x_\beta}\right)$$

$$= \sum_{\mu\nu} \sqrt{-g}\, \frac{\partial g_{\mu\nu}}{\partial x_\sigma} \left\{ \sum_{\alpha\beta} \frac{1}{\sqrt{-g}} \frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{-g}\, \frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) - \sum_{\alpha\beta\tau\rho} \gamma_{\alpha\beta}\, g_{\tau\rho} \frac{\partial \gamma_{\mu\tau}}{\partial x_\alpha} \frac{\partial \gamma_{\nu\rho}}{\partial x_\beta} \right.$$

$$\left. + \frac{1}{2}\sum_{\alpha\beta\tau\rho} \gamma_{\alpha\mu}\gamma_{\beta\nu} \frac{\partial g_{\tau\rho}}{\partial x_\alpha} \frac{\partial \gamma_{\tau\rho}}{\partial x_\beta} - \frac{1}{4}\sum_{\alpha\beta\tau\rho} \gamma_{\mu\nu}\gamma_{\alpha\beta} \frac{\partial g_{\tau\rho}}{\partial x_\alpha} \frac{\partial \gamma_{\tau\rho}}{\partial x_\beta} \right\} \tag{7.3b}$$

From comparison with equation (7.3a) Einstein was able to choose the term in the curly brackets on the right hand side as $\Gamma_{\mu\nu}$ and thus complete his derivation of the "Entwurf..." field equations. Also he was able to define a stress-energy tensor for the gravitational field and show that it entered into a conservation law of the form of equation (7.1a). Finally he could write his field equations in a form which indicated that the source of the gravitational field was the sum of the gravitational and non-gravitational stress-energy tensors. This confirmed that his theory was consistent with the conservation and democracy of energy-momentum.

But the "Entwurf..." field equations were not generally covariant and Einstein believed that his method had shown that they were the only possible second order field equations. This mistake did not arise from the

19. "eindeutig bestimmt". Ibid., p.237.

general method he had adopted to derive the field equations. His method was basically sound. I now demonstrate this by showing how a slightly modified version of this method can lead to the recovery of the 1915 generally covariant field equations in an argument that turns out to be very close to those used in more modern texts.

Using a more modern notation, we assume the field equations to have the form $G^{\mu\nu} = \varkappa T^{\mu\nu}$
where $G^{\mu\nu}$ corresponds to Einstein's $\Gamma_{\mu\nu}$. Instead of using Einstein's equation (7.1c) as a statement of the conservation law, we use the more familiar $T^{\mu\nu}{}_{;\nu} = 0$ (7.1d)
Substituting with the field equations, we conclude the existence of a set of identities $G^{\mu\nu}{}_{;\nu} \equiv 0$
constructed solely out of the metric tensor and its derivatives. Unlike Einstein we can retain the requirement of general covariance. Then the reader will immediately recognise these identities as the contracted Bianchi identities $\left(R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R\right)_{;\nu} \equiv 0$ (7.3c)
where $R^{\mu\nu}$ is the Ricci tensor and R the Riemann curvature scalar. Use of these identities rather than Grossmann's enables us to recover immediately Einstein's 1915 generally covariant field equation

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R = \varkappa T^{\mu\nu} \qquad (7.2b)$$

Thus Einstein's method seems quite capable of delivering his final generally covariant field equations, provided the appropriate identities (7.3c) are used. Recognition of the significance of these identities for Einstein's theory did not appear until well after Einstein's discovery of his 1915 field equations.[20] However, we cannot simply use the lack of awareness of these identities to explain Einstein's failure to discover generally covariant field equations in 1913. For built into the framework of the "Entwurf..." paper was an assumption which would have made the recovery of Einstein's later generally covariant field equations impossible and led him to believe that Grossmann's identities and the resulting non-generally covariant field equations were unique.

Einstein expected $\Gamma_{\mu\nu}$ to be a generalisation of the Laplacian of the scalar gravitational potential of Newtonian theory, $\Delta\varphi$. A natural candidate for this was the expression

$$\sum_{\alpha\beta} \frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) \qquad \text{(a)}$$

for he could argue that this term reduced to the expression

$$-\left(\frac{\partial^2\gamma_{\mu\nu}}{\partial x_1^{\,2}} + \frac{\partial^2\gamma_{\mu\nu}}{\partial x_2^{\,2}} + \frac{\partial^2\gamma_{\mu\nu}}{\partial x_3^{\,2}} - \frac{1}{c^2}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_4^{\,2}}\right)$$

in the case of a weak gravitational field in whose presence spacetime has almost the same structure as in special relativity. (x_1, x_2 and x_3 are Cartesian space coordinates and x_4 is the time coordinate. Therefore we recognise this latter expression to be the d'Alembertian of $\gamma_{\mu\nu}$, $\Box\gamma_{\mu\nu}$) From this Einstein was able to conclude:

> "If the field is static and only g_{44} variable, then we come to the case of Newtonian gravitation theory, if we put the constructed expression up to a constant for the quantity $\Gamma_{\mu\nu}$."[21]

---

20. J. Mehra, Einstein, Hilbert, and the Theory of Gravitation (Dordrecht: Riedel, 1974), pp.49-50 and Note 242, p.78.

21. "Ist das Feld ein statisches und nur g_{44} variabel, so kommen wir also auf den Fall der Newtonschen Gravitationstheorie, falls wir den gebildeten Ausdruck bis auf eine Konstante für die Grösse $\Gamma_{\mu\nu}$ setzen." Einstein, "Entwurf...", p.235.

He continued to note that the expression (a) could not be regarded to be the unique generalisation of $\Delta\varphi$ of Newtonian theory, for further terms which would vanish in first approximation in the case of a weak field could be added to it. As an example of such terms he mentioned

$$\sum_{\alpha\beta} \frac{\partial g_{\alpha\beta}}{\partial x_\mu} \frac{\partial \gamma_{\alpha\beta}}{\partial x_\nu}$$

Einstein made his assumptions on the form of $\Gamma_{\mu\nu}$ clearest when he summarised the form expected for the identities which I have numbered (7.3a):

> "The identity sought is therefore of the following form:
>
> Sum of differential quotients
>
> $$= \frac{1}{2}\sum_{\mu\nu} \sqrt{-g}\, \frac{\partial g_{\mu\nu}}{\partial x_\sigma} \left\{ \sum_{\alpha\beta} \frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta} \frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) \right.$$
>
> $$\left. + \text{ further terms which fall away with the formation of the first approximation} \right\}"$$ [22]

The term which represents $\Gamma_{\mu\nu}$ is in the curly brackets on the right hand side of the equation. Thus Einstein assumed $\Gamma_{\mu\nu}$ to have the form

$$\Gamma_{\mu\nu} = \sum_{\alpha\beta} \frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta} \frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) + \begin{array}{l}\text{terms which vanish to}\\ \text{a first approximation}\\ \text{in a weak field}\end{array} \qquad (7.4)$$

and that it must have this form if the field equations are to yield the correct Newtonian Limit.

Clearly this assumption is over-restrictive, for Einstein's 1915 generally covariant field equations do not satisfy this condition but still yield the correct Newtonian limit.

---

22. "Die gesuchte identische Gleichung ist also von folgender Gestalt: Summe von Differentialquotienten = [expression as above] + weitere Glieder, die bei Bildung der ersten Annäherung wegfallen." Einstein, "Entwurf...", p.237.

Finally this assumption guaranteed that Einstein would fail to recover generally covariant field equations from his otherwise sound "Entwurf..." method. For Grossmann addressed himself to the problem of constructing a set of identities of the form of (7.3a) in which $\Gamma_{\mu\nu}$ takes on the form given by equation (7.4). If the method were to yield the final generally covariant field equations, then Einstein would have had to work from a far more complex set of identi- tities. That is a restatement of the contracted Bianchi identities in the form of equation (7.3a). In the more usual modern notation these identities would read

$$\left(\sqrt{-g}\, t^{\nu}_{\sigma\,(GR)}\right)_{,\nu} = \tfrac{1}{2}\sqrt{-g}\; g_{\mu\nu,\sigma} \left\{ R^{\mu\nu} - \tfrac{1}{2} g^{\mu\nu} R \right\} \tag{7.3d}$$

where $t^{\nu}_{\sigma\,(GR)}$ is the gravitational stress-energy tensor of the final theory.[23] Now the term corresponding to $\Gamma_{\mu\nu}$ in these equations contains many more second derivatives of the metric tensor which do not vanish to first approximation in a weak field than is allowed by equation (7.4). Thus Grossmann would not have sought to construct identi- ties such as these, nor would Einstein have used them if he had.

This also provides a simple explanation of Einstein and Grossmann's rejection of the Ricci tensor as the basis of the field equations. All they would need do is write out the Ricci tensor explicitly in terms of the derivatives of the metric tensor to see that it did not have the form required by equation (7.4). As we have seen, this was the form Einstein expected the quantity $\Gamma_{\mu\nu}$ to have if the field equations

---

23. As given, for example, in P.A.M. Dirac, General Theory of Relativity (New York: John Wiley, 1975), p.63.

were to reduce to the correct Newtonian limit in the case of a weak static field.

Earlier, in Section 5.5, I described Stachel's explanation of this episode. It will be recalled that he suggested that the rejection of the Ricci tensor did not result from what is an essentially mathematical error - Einstein and Grossmann's failure to realise that the correct Newtonian limit could indeed be recovered from field equations based on the Ricci tensor. Rather, he suggested that the rejection followed from Einstein's assumption that static gravitational fields are spatially flat. Calculation of the Ricci tensor for such a field shows that this tensor could not be the basis of field equations capable of dealing with the range of fields known to exist.

To begin with, the first explanation, rather than Stachel's, seems to account most plausibly for the immediate reason for the rejection of the Ricci tensor. In the "Entwurf..." paper, Einstein tells us quite clearly which conditions he believed had to be satisfied if the field equations were to yield the correct Newtonian limit. This is embodied in equation (7.4). That the Ricci tensor did not satisfy this condition would have been apparent from a simple inspection of its form. Presumably this would have been seen by Einstein or Grossmann well before any calculation of the values of the components of the tensor would have been carried out for the special case of a static field.

However, this explanation is incomplete, as it now stands, for only describes the final step of a more complex interaction of concepts and ideas. In particular, Einstein's assumptions about the requirements which had to be fulfilled if a correct Newtonian limit was to be recovered were closely linked with his belief that static fields must be spatially flat.

The assumption that $\Gamma_{\mu\nu}$ must satisfy equation (7.4) meant that the field equations must reduce to the transparent

$$\Box g_{\mu\nu} = \varkappa T_{\mu\nu} \qquad (7.2c)$$

for the case of a weak field.* This, of course, was the immediate purpose of imposing the condition. If the source mass distribution is static, then these equations reduce to the set of equations

$$\Delta g_{44} = \varkappa \rho c^2 \qquad (7.2d)$$

where ρ is the rest density of the source mass distribution

and $$\Delta g_{\mu\nu} = 0 \qquad (7.2e)$$

if $\mu \neq 4$ and $\nu \neq 4$. For the boundary conditions of this special case, equation (7.2e) solves to yield the result that the $g_{\mu\nu}$ are constant, for the case in which $\mu \neq 4$ and $\nu \neq 4$. This, of course, entails the result that the spatial cross-section of spacetime is flat. Also it can be seen from equation (7.2d) that the more general equation (7.2c) reduces to the case of Newtonian gravitation theory, with g_{44} standing for the Newtonian gravitation potential, in this case of a static source mass distribution.

Thus, in short, we can say that the imposition of condition (7.4) on the form of $\Gamma_{\mu\nu}$ entails the result that the field equations yield a weak field limit in which only g_{44} is variable and can stand for the Newtonian gravitational potential and which is spatially flat. We know Einstein was aware of this in 1913 for he derived exactly

---

* Here I continue to use Einstein's 1913 notation.

the results presented above in equations (7.2d) and (7.2e) and the constancy of the relevant components of $g_{\mu\nu}$ in his 1913 Vienna address.[24]

However, in spite of this, in the "Entwurf..." paper, when he sought to demonstrate that the imposition of condition (7.4) was sufficient to ensure generation of the correct Newtonian limit, Einstein completed his argument by introducing the assumption that static gravitational fields are spatially flat as an independent assumption. This can be seen quite clearly in my account of his argument earlier in this section. (I quoted Einstein's remarks at the crucial point. There he wrote: "If the field is static and only g_{44} variable...".) This enabled Einstein to reduce the second rank tensor expression $\Box\gamma_{\mu\nu}$ to the scalar expression $\Delta\gamma_{44}$, a reduction which he presumably felt necessary to demonstrate the adequate recovery of a Newtonian limit. The assumption that the field is static in itself would enable Einstein only to drop the time derivative terms of $\Box\gamma_{\mu\nu}$ yielding $\Delta\gamma_{\mu\nu}$. The additional assumption that such fields are spatially flat ("...only g_{44} variable..") enabled him to set to zero all but the γ_{44} components from this term.

Finally, it will be recalled that, earlier in Section 5.5, I argued that this assumption of the spatial flatness of static fields had been imported into the "Entwurf..." paper directly from Einstein's earlier scalar theory of gravitation.

This enables us to put together a more complete picture of the considerations underlying the rejection of the Ricci tensor. It resulted from the adoption of two assumptions, which originally arose

24. Einstein, "Zum gegenwärtigen Stande...", p.1259.

independently. The first came from Einstein's earlier scalar gravitation theory and was that static fields are spatially flat. The second came from Einstein's search for acceptable field equations, in which he made the natural assumption that such equations must reduce to the simple form of equation (7.2c). In 1913, Einstein would have found that both assumptions meshed together and reinforced one another in the manner outlined above. This would have been most encouraging, for it is just the sort of unexpected agreement one hopes will emerge if one's developing theory is internally consistent. However, neither assumption holds within the final theory and thus the understanding of gravitational fields which stemmed from their fusion made the use of the Ricci tensor in the field equations quite impossible.

7.4.2 The Principle of Equivalence and the 1912 Scalar Theory

In an earlier chapter (Section 4.6) I described briefly how Einstein had discovered that the original field equation for his 1912 scalar theory were inconsistent with the equality of action and reaction and how this had forced him to concede reluctantly that the equivalence of acceleration and gravitation did not hold globally even for the special case of uniform acceleration and homogeneous gravititational fields. In this section, I will characterise this episode as a collision with the constraint envelope and show how Einstein was prepared to allow modification to the hypotheses of Group B rather than admit any compromise of the contents of the constraint envelope.

It will be recalled that in his 1912 theory Einstein had concluded from the global equivalence of uniform acceleration and homogeneous gravitational fields that the gravitational potential of a homogeneous field was given by the speed of light, c, which varied linearly with distance in the direction of the field. This suggested a source-free field equation which he wrote as

$$\Delta c = \frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} + \frac{\partial^2 c}{\partial z^2} = 0$$

This equation generalised to the field equation

$$\Delta c = kc\sigma \tag{7.5a}$$

where k is a constant and σ is the self measured rest density of matter. c must appear as a factor on the right hand side in order to ensure that c is determined by the equation only up to a constant factor.[25]

In the second of his two papers presenting his 1912 theory, Einstein displayed the contradiction which followed from this field equation.[26] Within the theory, the force density on a static mass distribution was written by Einstein as

$$\mathfrak{F} = -\sigma \,\mathrm{grad}\, c \tag{7.6}$$

He then calculated the net force acting on a system of masses connected by a rigid massless frame in a space in which c is constant at spatial infinity. With substitution from the field equation this was

$$\int \mathfrak{F}\, d\tau = -\int \sigma \,\mathrm{grad}\, c \, d\tau = -\frac{1}{k}\int \frac{\Delta c}{c} \,\mathrm{grad}\, c \, d\tau \tag{7.7}$$

where $d\tau$ represents a spatial volume element.

Now in general, this expression does not vanish. This implies that there is a net residual force on the system and that it will set itself

---

25. A. Einstein, "Lichtgeschwindigkeit und Statik des Gravitationsfeldes", Annalen der Physik, 38 (1912), p.360; A. Einstein, "Zur Theorie des statischen Gravitationsfeldes", Annalen der Physik, 38 (1912), p.452.

26. Ibid. The rest of this section deals with the contents of the remainder of this second paper of Einstein's.

in motion. This, Einstein concluded, was a violation of the equality of action and reaction.

In the terms of this chapter, we can recognise this as a violation of a constraint of Group A, the conservation of energy and momentum. Alternatively we can regard the contradiction as arising from the inability of the theory to provide a three-dimensional stress tensor for the gravitational field. For, by definition, the integrand in equation (7.7) is the divergence of this tensor and, as a result, the integral should vanish as a surface integral. Thus we would expect to be able to rewrite the integrand $-\frac{1}{k}\frac{\Delta c}{c}\,\mathrm{grad}\,c$ as the divergence of an expression which could then be seen to be the gravitational field stress tensor. However, no such rewriting seems possible. Therefore it seems that such a tensor cannot be constructed, the integral does not vanish and an anomalous net force on the system of masses remains.

Having recognised the collision with the requirements of the constraint envelope, Einstein then turned to examine possible resolutions. First he considered the possibility that gravitational fields might act not only on the masses of the system in question, but also on the stressed but massless frame connecting them. However, this possibility had to be immediately rejected for it contradicted one of the implications of the democracy of energy-momentum, the equality of passive gravitational and inertial mass.

Einstein considered some light energy contained within a massless, mirrored box. Now the gravitational mass of the system would be made up from the energy of the light as well as a contribution from the stresses in the walls of the box due to the pressure of the radiation. But from special relativity, he continued, we know that the inertial mass of the entire system comes only from the energy of the light. The stresses in the box walls make no contribution. Thus the two quantities could only be equal if gravitational fields are assumed not to act on stressed but massless frames. A similar conclusion followed from the consideration of the analogous case of particles in rapid motion colliding elastically with the walls of a container. (We shall see in the next chapter how Einstein returned to these thought experiments in the following year and, with an appropriate choice of source term, concluded that it was possible for gravitation to be thought of as acting on stresses, without violating the requirement of the equality of inertial and gravitational mass.)

Then Einstein considered another possibility, the modification of the force law (7.6). Once again Einstein found that the requirements of the constraint envelope would not allow this. For modification of this law would require modification of the equations of motion of a point mass within the theory, since these were the equations on which the force law rested. Einstein painstakingly argued that the requirement of the special relativistic limit, the equality of inertial and gravitational mass and dimensional considerations only allowed the modification of these equations through the insertion of multiplicative constants, consisting of c raised to an arbitrary power. He then considered the term

$$\mathfrak{R}_{xa}\dot{x} + \mathfrak{R}_{ya}\dot{y} + \mathfrak{R}_{za}\dot{z}$$

where $\mathfrak{R}_{xa}, \mathfrak{R}_{ya}$ and $\mathfrak{R}_{za}$ are the components of the non-gravitational force acting on the point mass and $\dot{x}, \dot{y}$ and $\dot{z}$ are the components of its velocity. We recoqnise this term to be the rate at which these forces transfer energy to the mass point. Einstein concluded that this term could only be written as a differential quotient with time if the new multiplicative constants were adjusted to yield the case of his existing scalar theory. (It is clear from the requirement of the conservation of energy that this term must be capable of being written in this form if an energy balance is to be constructed.) Thus Einstein was forced back to the equation of motion and the force law of his original theory.

As a result Einstein resolved himself to the final possibility, a modification of the field equation . Manipulation of the integrand of equation (7.7) readily suggested which term had to be added to the field equation in order that the modified integral $\int \mathfrak{F} d\tau$ should now vanish. The modified field equation was

$$\Delta c = k \left\{ c\sigma + \frac{1}{2k} \frac{\operatorname{grad}^2 c}{c} \right\} \tag{7.5b}$$

and could be rewritten in the simpler form

$$\Delta(\sqrt{c}) = \frac{k}{2} \sqrt{c}\, \sigma \tag{7.5c}$$

This resolved Einstein's immediate difficulties. Using the modified field equation , there was calculated to be no net force on the mass system considered earlier. Also he was able to define a stress tensor for the gravitational field. And there was an unexpected payoff. Einstein was able to show that the extra term appearing on the right hand side of the field equation (7.5b) was equal to the energy density of the gravitational field. This once again confirmed the satisfaction of the requirements of the democracy of energy-momentum within the theory. For here was visible proof that all energy, including the energy of the gravitational field itself, contributed equally as an active source of the gravitational field.

But, Einstein tells us, he turned to this last possibility with difficulty, for it involved sacrificing the global equivalence of uniform acceleration and homogeneous gravitational fields. This global equivalence entailed the linear dependence of c with distance in the direction of the homogeneous field. This had been entailed by the first field equation (7.5a). But it was no longer entailed by the modified equation. Inspection of this equation in the form of (7.5c) shows that the new equation entails a linear dependence of $\sqrt{c}$ with distance in the direction of a homogeneous gravitational field. As a result Einstein was forced to limit the equivalence to regions of space of indefinitely small extent. Fortunately he was able to note that this did not materially affect the remainder of his theory.

7.4.3 Review

In retrospect there are two points which I would like to stress about these episodes:

First comes a point about the dynamics of Einstein's research programme. In both episodes we see quite clearly the primacy of the constraints of Group A over the hypotheses of Group B. For rather than allow any compromise of the contents of Group A, Einstein was prepared to admit significant modifications to the hypotheses of Group B - he was willing to give up the requirement of general covariance and the global equivalence of uniform acceleration and homogeneous gravitational fields.

Second we see the power of the constraint envelope to lead Einstein to a well determined result. In the case of the "Entwurf..." field equations, the requirements of the conservation of energy-momentum and of the Newtonian limit forced Einstein to a specific set of field equations. It became clear later that an overrestrictive interpretation of the requirement of a Newtonian limit led to what amounted to an overdetermination of the field equations. However, I have argued earlier in this chapter that a more careful following through of the implications of the constraint envelope leads to a very restricted choice of field equations, which includes those of the final theory.

In the second episode, we see once again how the constraint envelope forced Einstein to a quite specific result - in this case the limitation of the equivalence of acceleration and gravitation to local regions. In particular, Einstein's account of the episode gives us an interesting insight into the processes at work in Einstein's discovery of this result. For it seems as if we are following Einstein from his first discovery of the violation of the requirements of the constraint enevelope through each of the various resolutions which he attempted. As we follow him in this, we see how the direction of his path is decisively controlled by the requirements of the constraints of Group A, most of which are invoked explicitly - the equality of inertial and gravitational mass, the conservation of energy and the special relativistic limit. Finally, he is forced reluctantly to a specific modified field equation and the weakening of the equivalence of acceleration and gravitation.

7.5 Conclusion

Einstein was guided towards his discovery of the general theory of relativity by an organic set of heuristics at whose heart lay the principle of equivalence and the concept of the general relativity of motion. The organic nature of this set, however, poses a serious problem for the historian who is interested in the rational reconstruction of the origins of the general theory of relativity. For the set grew and changed continuously with the development of the theory until key aspects of its original contents were contradicted by the theory which finally emerged.

I have argued that it is still possible to understand the emergence of the theory as the outcome of a single rational process. For beneath all the changing complexities lay the challenge of a single task, that of constructing a relativistically acceptable gravitation theory. This formed a stable and well-defined constraint enevelope within which Einstein worked from the very beginning and whose requirements he refused to compromise. We can look upon the heuristics mentioned above as a force which drove Einstein onwards in his search for a general theory of relativity. But it was the constraint envelope which guaranteed that Einstein would finally converge on the general theory of relativity or one much like it, as long as he teased out its implications to the full.

Furthermore, it was the uncompromisable stability of this constraint enevelope which ensured the integrity and continuity of Einstein's research programme, in spite of the instability of the driving heuristics. Indeed, I have even argued that on a number of occasions changes in these heuristics were forced on Einstein by a contradiction between them and the constraint envelope.

In arguing for such an account of the origins of the general theory of relativity, I wish in no way to diminish the magnitude of Einstein's great achievement. If anything my analysis stresses how remarkable Einstein's achievement was. For it was he alone who foresaw the far reaching implications of an apparently innocuous problem, the construction of a relativistically acceptable gravitation theory. He was able to anticipate important concepts and results which would follow from this problem well before they were forced upon him. Finally, he was able to fuse a wide range of disparate physical, philosophical and methodological considerations into an organic unity out of which, virtually single-handedly, he was able to precipitate and consummate one of the most important revolutions in modern physics.

CHAPTER 8

NORDSTRÖM'S THEORY OF GRAVITATION

8. NORDSTRÖM'S THEORY OF GRAVITATION

In the last chapter, I reviewed the prospects for gravitation theory within special relativity. I described how the application of natural consistency requirements led to the rejection of a vector theory and the construction of a complex second rank tensor theory, which turned out to be formally identical to the general theory of relativity. In this chapter, I return to the examination of the prospects of a scalar gravitation theory within special relativity. As we shall see there is a strong similarity between this case and that of the second rank tensor theory. For once again we find that the imposition of natural consistency requirements leads to the conclusion that rods and clocks do not behave in accord with the requirements of special relativity - in short, gravitation is again seen to burst out of special relativity. Indeed the theory that results is formally strongly analogous to the general theory of relativity. It can be written as a generally covariant theory, within a Riemannian spacetime with a field equation based on the Riemann curvature tensor.

This result takes on special significance here, for it adds further support to my earlier argument that Einstein was constrained to burst out of special relativity and arrive at a theory much like the general theory of relativity simply by virtue of the fact that he was seeking a relativistically acceptable gravitation theory.

The teasing out of the prospects of a scalar gravitation theory within special relativity turns out to be an episode of historical interest to us as well. For it was a process which was carried out by a colleague of Einstein's, the physicist Gunnar Nordström, at the same time as Einstein was developing his general theory of relativity, and with Einstein's critical assistance.

In the case of his general theory of relativity, Einstein was able to anticipate results crucial to the character of his final theory well before they could have been forced upon him by the constraint envelope For example, already in 1907 he had concluded that gravitation was to burst out of special relativity but that an acceptable account of gravitation could be found by incorporating it into an extended theory of space and time. This tends to mask the significance of the constraint enevelope for the emergence of Einstein's theory. This is not the case with Nordström's theory. We shall see that constraints similar to those outlined in the previous chapter played a central and continuing role in the shaping of the final form of the theory. I now turn to the task of tracing this process.

8.1 Nordström's First Theory

Nordström introduced his first paper in 1912 on his gravitation theory with a clear statement of its purpose:

> "The hypothesis of Einstein, that the speed of light c depends on the gravitational potential, leads to considerable difficulties for the principle of relativity, as emerges from the discussion between Einstein and Abraham. Therefore the question arises whether it is not possible to replace Einstein's hypothesis with another, which leaves c constant and still adapts the theory of gravitation to the principle of relativity in such a way that gravitational and inertial masses become equal. I believe I have found such a hypothesis and want to present it in the following."[1]

---

1. "Die Hypothese Einsteins, dass die Geschwindigkeit c des Lichts vom Gravitationspotential abhänge, führt, wie aus der Diskussion zwischen Einstein und Abraham hervorgeht, zu erheblichen Schwierigkeiten für das Relativitätsprinzip. Es fragt sich deshalb, ob es nicht möglich ist, die Einsteinsche Hypothese durch eine andere zu erstzen, die c konstant lässt und doch die Theorie der Gravitation in solcher Weise dem Relativitätsprinzip anpasst, dass schwere und träge Masse gleich werden. Ein solche Hypothese glaube ich gefunden zu haben und will im folgenden dieselbe darstellen." G. Nordström, "Relativitätsprinzip und Gravitation", Physikalische Zeitschrift, 13 (1912), p.1126.

In other words, Nordström sought to construct a gravitation theory within special relativity, one that we would now describe as Lorentz covariant. The theory which he proceeded to describe in his two consecutive papers in 1912 is formally identical with the Lorentz covariant scalar theory outlined in Section (3.2) here. [2] Its field equation was *

$$\Box^2 \varphi = -g\nu \tag{8.1a}$$

where ν is rest density of matter and g, the "gravitation factor", converts inertial mass to gravitational mass, thus ensuring that the right hand side of this field equation represents a source density. Rational units are used so that the gravitational constant, which usually appears here, is equal to unity. The force equation was given by

$$\mathcal{R}^3_\mu = g\nu\varphi_{,\mu} \tag{8.2a}$$

where $\mathcal{R}^3_\mu$ represents the proper density of four-force.

In the first of the two papers, Nordström was content to describe

---

* In this chapter I continue to use the compact notation of Chapter 3, rather than that used by Nordström and Einstein, which often requires all components of a tensor equation to be written out in full. This makes the presentation of the contents of the equations and their derivation more compact and transparent. Where possible I try to keep close to the original in the choice of symbols. Where Nordström and Einstein use a metric equivalent to $ds^2=du^2+dx^2+dy^2+dz^2$, with u=ict, I use $ds^2=c^2dt^2-dx^2-dy^2-dz^2$ and retain the distinction between covariant and contravariant tensors. This accounts for differences in sign in equations such as (8.2).

2. Ibid. and G. Nordström, "Träge und schwere Masse in der Relativitätsmechanik", Annalen der Physik, 40 (1912), pp.856-78. For a contemporary review of Nordström's theories, see M. von Laue, "Die Nordströmsche Gravitationstheorie", Jahrbuch der Radioaktivität und Elektronik, 14 (1917), pp.263-313.

the source density ν as the "rest density of matter".[3] In the second, however, he engaged in a lengthy analysis of the concept of mass, in the context of the relativitistic mechanics of deformable bodies, and ultimately decided on a definition of ν which set it equal to $1/c^2$ (rest energy density).[4]

Nordström recognised the implications of the identity (3.5) and from it concluded a result analogous to equation (3.9). Rest mass m must vary with the gravitational potential according to

$$m = m_0 \exp \frac{g\varphi}{c^2} \tag{8.3a}$$

where m_0 is a constant characteristic of the mass in question. Nordström acknowledged the possibility of retaining the independence of m from the gravitational potential by modifying the force law in a manner analogous to that described in Section 3.2 here. He noted that, whilst each gave a different force law and expression for mass, both gave exactly the same laws for the motion of point masses in gravitational fields. He also presented a straightforward stress-energy tensor for the gravitational field and an analysis of the physical meaning of its components.

Finally there was some discussion of the trajectories of freely falling bodies in the theory. The theory entailed that the downwards acceleration of a mass point in a homogeneous gravitational field directed in the z-direction was given by

$$\frac{dv_z}{dt} = -\left(1 - \frac{v^2}{c^2}\right) g \frac{\partial\varphi}{\partial z} \tag{8.4}$$

---

3. "Ruhdichte der Materie". Nordström, "Relativitätsprinzip...", p.1126.

4. See especially Nordström, "Träge und schwere Masse...", pp.866-8. See also A.L. Harvey, "Brief Review of Lorentz Covariant Theories of Gravitation", American Journal of Physics, 33 (1965), p.455.

where v is the velocity of the mass.[5] That is, its downwards acceleration varies with the sideways velocity of the mass. This was exactly the result which had led Einstein to reject such a scalar theory and which I argued in Chapter 3 will inevitably appear in all such theories. Einstein communicated his objection to Nordström and it was acknowledged in a postscript to the first of Nordström's papers on his theory:

> "From a postal communication from Herr Prof. Dr. A. Einstein I learn that he has occupied himself already earlier with the possibility used by me above of treating gravitational phenomena in a simpler way, but that he had come to the conviction that the consequences of such a theory cannot correspond with reality. He shows in a simple example that a rotating system will receive a smaller acceleration in a gravitational field than one not rotating."[6]

In his discussion immediately following, Nordström did not seem too concerned by this result, for he noted that the difference in accelerations was too small to be detected in experience. However, in the second paper on his theory, which he published very soon after, he recognised a serious ramification of this feature of his theory. For he now saw that the molecular motion of a falling body should be seen to have an influence on its acceleration when freely falling according to this theory. He mentioned the possibility of incorporating this

---

5. Nordström, "Träge und schwere Masse...", p.877-8. See also M. Behacker, "Der freie Fall und die Planetenbewegung in Nordströms Gravitationstheorie", Physikalische Zeitschrift, 14 (1913), pp.989-92.

6. "Aus einer brieflichen Mitteilung von Herrn. Prof. Dr. A. Einstein erfahre ich, dass er sich bereits früher mit der von mir oben benutzten Möglichkeit befasst hat, die Gravitationserscheinungen in einfacher Weise zu behandeln, dass er aber zu der Überzeugung gekommen ist, dass die Konsequenzen einer solchen Theorie die Wirklichkeit nicht entsprechen können. Er zeigt an einem einfachen Beispiel, dass nach dieser Theorie ein rotierendes System im Schwerkraftfelde eine kleinere Beschleunigung erhalten wird als ein nichtrotierendes." Nordström, "Relativitätsprinzip...", p.1129.

explicitly into his theory by allowing the constant g to vary with the molecular motion of the body in question, but left the possibility unexplored for the time being.[7]

8.2 Einstein's Challenge to Scalar Gravitation Theories

At the conclusion of the physical part of his 1913 "Entwurf..." paper, Einstein returned to the question of the possibility of a scalar theory of gravitation. He made his motives explicit. His new tensor theory was markedly more complex than the more familiar scalar theories and he felt the need to justify why he had chosen this more complex option. Einstein was quite correct in anticipating criticism on this point. In the following year, Abraham, for one, criticised Einstein's theory on exactly this ground, invoking, in particular, Mach's concept of the economy of thought.[8] This was a penetrating rhetorical device in view of Einstein's repeatedly expressed hope that his new theory would realise the requirements of Mach's critique of Newtonian mechanics.

Because of the profound influence that his observations would have on the development of Nordström's theory, I quote them at length:

> " § 7. Can Gravitational Fields be Reduced to a Scalar?
>
> With the undeniable complexity of the theory advocated here, we must ask ourselves seriously whether the understanding exclusively advocated until now, in which gravitational fields are reduced to a scalar Φ , might be the only

---

7. Nordström, "Träge und schwere Masse...", p.878.

8. M. Abraham, "Neuere Gravitationstheorien", Jahrbuch der Radioaktivität und Elektronik, 11 (1914), p.520.

obvious and justified one. I will demonstrate briefly why I believe we must answer this question in the negative.

There is a path for characterisation of gravitational fields, which is quite analogous to that struck in the preceding. One begins with the equation of motion of a material point in Hamiltonian form

$$\delta\left\{\int \Phi\, ds\right\} = 0,$$

where ds is the line element of the usual theory of relativity and Φ is a scalar, and then continues quite analogously to the preceding, without having to abandon the usual theory of relativity.

Here also material processes of arbitrary kind are characterised through a stress-energy tensor $T_{\mu\nu}$. But, in this understanding, a scalar measures the interaction between gravitational fields and material processes. This scalar, which Herr Laue pointed out to me, can only be

$$\sum_{\mu} T_{\mu\mu} = P$$

which I want to name "Laue's scalar"...Then one can do justice also here, up to a certain degree, to the law of the equivalence of inertial and gravitational mass. For Herr Laue pointed out to me that for a closed system

$$\int P\, dV = \int T_{44}\, d\tau \; *$$

From here it is seen that also according to this understanding the total energy of a closed system determines its weight.

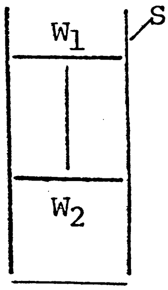
The weight of a system which is not closed, however, will depend on the orthogonal stresses T_{11} etc., to which the system is subjected. From this arise consequences which appear to me to be unacceptable, as shall be shown in the example of cavity radiation.

The scalar P disappears for radiation in a vacuum, as is well known. If the radiation is closed up in a massless mirrored box, then its walls experience tensions, which act so that the system - taken as a whole - has a gravitational mass $\int P\, d\tau$, which corresponds to the energy E of the radiation.

Now, however, instead of closing up the radiation in an empty box, I imagine it bounded

---

* This appears to be a misprint. "dV" should read "$d\tau$", where $d\tau$ is a volume element.



1. through the mirrored wall of a fixed pit S,
2. through two mirrored, moveable walls W_1 and W_2, which are solidly fixed to one another with a rod.

In this case the gravitational mass $\int P dt$ of the moveable system amounts to only a third of the value, which appeared when the box moved as a whole. Therefore one would need to expend only a third part of the work in raising the radiation against a gravitational field as in the case considered just now, in which the radiation is closed up in a box. This appears to me to be unacceptable.

Certainly I must admit that for me the most effective argument for the case that such a theory should be rejected rests on the conviction that relativity should not only hold for linear orthogonal subsitutions but for a much wider group of substitutions. But we are not yet entitled to make good this argument, since we are not in a position to find the (most general) group of substitutions which belong to our gravitation equations." [9]

9. "§ 7. Kann das Gravitationsfeld auf einen Skalar zurückgeführt werden?

Bei der unleugbaren Kompliziertheit der hier vertretenen Theorie der Gravitation müssen wir uns ernstlich fragen, ob nicht die bisher ausschliesslich vertretene Auffassung, nach welcher das Gravitationsfeld auf einen Skalar Φ zurückgeführt wird, die einzig naheliegende und berechtigte sei. Ich will kurz darlegen, warum wir diese Frage verneinen zu müssen glauben.

Es bietet sich bei Charakterisierung des Gravitationsfeldes durch einen Skalar ein Weg dar, welcher dem im Vorhergehenden eingeschlagenen ganz analog ist. Man setzt als Bewegungsgleichung des materiellen Punktes in Hamiltonscher Form an

$$\delta\left\{\int \Phi\, ds\right\} = 0,$$

wobei ds das vierdimensionale Linienelement der gewöhnlichen Relativitätstheorie und Φ ein Skalar ist, und geht dann ganz analog vor wie im Vorhergehenden, ohne die gewöhnliche Relativitätstheorie verlassen zu müssen.

Auch hier ist der materielle Vorgang beliebiger Art durch einen Spannungs-Energie-Tensor $T_{\mu\nu}$ charakterisiert. Aber es ist bei dieser Auffassung ein Skalar massgebend für die Wechselwirkung zwischen Gravitationsfeld und materiellem Vorgang. Dieser Skalar kann, worauf mich Herr Laue aufmerksam machte, nur

$$\sum_{\mu} T_{\mu\mu} = P$$

sein, den ich als den „Laueschen Skalar" bezeichnen will... Dann kann man dem Satz von der Äquivalenz der trägen und der schweren Masse auch hier bis zu einem gewissen Grade gerecht werden. Herr Laue wies mich nämlich darauf hin, dass für ein abgeschlosseness System

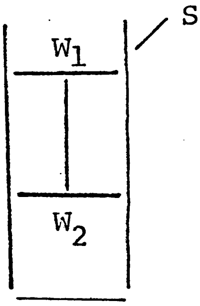
$$\int P\, dV = \int T_{44}\, d\tau$$

ist. Hieraus ersieht man, das für die Schwere eines abgeschlossenen Systems auch nach dieser Auffassung seine Gesamtenergie massgebend ist.

Die Schwere nicht abgeschlossener Systeme würde aber von den orthogonalen Spannungen T_{11} usw. abhängen, denen das System unterworfen ist. Daraus enstehen Konsequenzen, die mir unannehmbar erscheinen, wie an dem Beispiel der Hohlraumstrahlung gezeigt werden soll.

Für die Strahlung im Vakuum verschwindet bekanntlich der Skalar P. Ist die Strahlung in einem masselosen spiegelnden Kasten eingeschlossen, so erfahren deren Wände Zugspannungen, die bewirken, dass dem System, - als Ganzes genommen - eine schwere Masse $\int P\, d\tau$ zukommt, die der Energie E der Stahlung entspricht.

Statt nun aber die Stahlung in einen Hohlkasten einzuschliessen, denke ich mir dieselbe begrenzt



1. durch die spiegelnden Wände eines festangeordneten Schachtes S,

2. durch zwei vertikal verschiebbare spiegelnde Wände W_1 und W_2, welche durch einen Stab fest miteinander verbunden sind.

In diesem Falle beträgt die schwere Masse $\int P d\tau$ des beweglichen Systems nur den dritten Teil des Wertes, der bei einem als Ganzes beweglichen Kasten auftritt. man würde also zum Emporheben der Strahlung entgegen einem Schwerefelde nur den dritten Teil der Arbeit aufwenden müssen als in dem vorhin betrachteten Falle, dass die Strahlung in einem Kasten eingeschlossen ist. Dies erscheint mir unannehmbar.

Ich muss freilich zugeben, dass für mich das wirksamste Argument dafür, dass eine derartige Theorie zu verwerfen sei, auf der Überzeugung beruht, dass die Relativität nicht nur orthogonalen linearen Substitutionen gegenüber besteht, sondern einer viel weiteren Substitutionsgruppe gegenüber. Aber wir sind schon deshalb nicht berechtigt, dieses Argument geltend zu machen, weil wir nicht instande waren, die (allgemeinste) Substitutionsgruppe ausfindig zu machen, welche zu unseren Gravitationsgleichungen gehört."

A. Einstein and M. Grossmann, "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation", <u>Zeitschrift fur Mathematick und Physik</u>, <u>62</u>(1913), pp.243-4.

In this, there are three points of special interest. First is Einstein's suggestion that the simple Hamiltonian equation of motion given would be a natural starting point for a scalar gravitation theory, which Einstein himself would follow up later with success.

Second, there is his claim that the trace of the stress-energy tensor is the natural choice of the scalar quantity to measure gravitational mass. This claim needs little justification for modern readers, for this scalar is the one most readily constructed from the stress-energy tensor, which represents, in covariant form, the density of all matter in a given region.

However, as a brief justification of this choice, Einstein referred to the work of Laue, who had examined the properties of the general stress-energy tensor in some detail. In particular, he had shown that for a closed system in static equilibrium in its rest frame

$$\int T_{ij}\, dV = 0 \qquad\qquad i,j = 1,2,3$$

where dV represents the three dimensional volume element and the integration extends over the whole system. From this it followed that the integral of the trace of the matter tensor was given by the total energy of the system,

$$\int T^{\mu}_{\mu}\, dV = \int T^{0}_{0} + T^{1}_{1} + T^{2}_{2} + T^{3}_{3}\, dV = \int T^{0}_{0}\, dV \qquad (8.5)$$

thus suggesting that this trace did indeed represent a scalar measure of the density of matter. Whilst Einstein gave the trace Laue's name, Laue himself seems not to have discussed its significance explicitly in his publications, before Einstein's mention of it.[10]

10. M. von Laue, "Zur Dynamik der Relativitätstheorie", Annalen der Physik, 35 (1911), pp.539-42; M. von Laue, Die Relativitätstheorie Vol I (Braunschweig: Vieweg, 1952), pp.149-50.

With this choice of source term Einstein had at least partially resolved his original objection to Nordström's theory. With Nordström's original choice of source term, rest mass density, the kinetic energy of rotation of a rotating body did not contribute to its gravitational mass. It did, however, contribute to its inertial mass. Hence, it followed that a rotating body fell slower than a non-rotating one. A similar argument applied to the more important case of a body composed of moving molecules. The argument no longer held if the trace of the stress-energy tensor is taken as the source term. For, using Laue's result, the gravitational mass of the system in each case will be exactly proportional to its energy or, in other words, its inertial mass.* Thus internal motions will not affect the rate of fall of a body. Or, in general, the overall rate of fall of a closed system will not depend on its internal state.

Third, from this apparently satisfactory choice of source terms, the most serious of results was to come, as Einstein proceeded to show. For it entailed that stresses in a body must contribute to its gravitational mass. Or, more figuratively, that "stresses have weight". This seemed to lead directly to a violation of the law of conservation of energy.

To show this Einstein described two ways of raising light energy, which is trapped inside massless containers with mirrored walls. Now the weight of such systems comes only from the stresses in the members of the container resulting from radiation pressure, since the trace of the stress-energy tensor vanishes for free radiation. Therefore in

---

* We are used to associate the trace of the stress-energy tensor with rest mass density. In the two cases considered above, however, it will be given by the rest mass density plus an additional term due to stresses set up by the centrifugal forces or by those needed to constrain the moving molecules to a closed system. This additional term, according to Laue's result, exactly makes up the difference between rest and total energy density.

the first case described by Einstein, the work required to lift the radiation would be just that required to lift the six stressed walls of the enclosing box. In the second case, the work required to lift the same amount of radiation would be just that required to lift the stressed rod connecting walls W_1 and W_2. Thus it follows that three times as much energy is needed in the first case as in the second.

With a little thought, it can be seen that this is sufficient to enable a cycle to be set up in which radiation is lowered in a closed box and raised in a shaft - a cycle which leads to a net gain in energy. Presumably such implications are what Einstein found unacceptable and thus demonstrated the impossibility of a scalar gravitation theory.

8.3 Nordström's Second Theory

Later in 1913, following the "Entwurf..." paper, Nordström described a modified version of his earlier theory, with which he hoped to meet the objections that Einstein had raised.[11] The starting point of the modification was the explicit attempt to incoporate the equality of inertial and gravitational mass into his theory. This would enable the introduction of the trace of the stress-energy tensor in the place of ν . He now recognised that his earlier definition of ν was unsatisfactory.[12]

In the modified form, the basic field and force equations (8.1a) and (8.2a) remained unaltered. However, the "gravitation factor" g now was allowed to vary as a function of the gravitational potential φ .

---

11. G. Nordström, "Zur Theorie der Gravitation vom Standpunkt des Relativitätsprinzips", Annalen der Physik, 42 (1913), pp.533-54. See also G. Nordström, "Über den Energiesatz in der Gravitationstheorie", Physikalische Zeitschrift, 15 (1914), pp.375-80.

12. Ibid., p.533.

In order to enable the invocation of the equality of inertial and gravitational mass, Nordström considered the simplified case of a system of masses, which are allowed to execute motions in a stationary configuration in a static gravitation field. The gravitational potential takes on the background value φ_a at infinity.

A direct application of Laue's result (8.5) gave the total energy of the system E_o as

$$E_o = \int D + L + G \, dV$$

where D,L and G are the traces of the stress-energy tensors of normal matter, the electromagnetic field and the gravitational field and the integration with the volume element dV extends over the entire volume of the system. Now L vanishes and, making use of lengthy but unproblematical manipulation of the expression for G, which involved the use of the field equation, Nordström concluded that

$$E_o = mc^2 = \int \left(D - (\varphi - \varphi_a)\, g(\varphi)\, \nu \right) dV \qquad (8.6)$$

As indicated, due to the inertia of energy, this is equal to mc^2, where m is the inertial mass of the system, including its gravitational field energy.

From the definition of the gravitation factor $g(\varphi)$, Nordström concluded that the gravitational mass of the system was given by

$$M_g = \int g(\varphi)\, \nu \, dV = g(\varphi_a)\, m \qquad (8.7)$$

The second equality amounts to the invocation of the proportionality of inertial and gravitational mass, with $g(\varphi)$ the constant of proportionality.

Combining equations (8.6) and (8.7), Nordström recovered the result $D = g(\varphi)\nu\left\{\varphi - \varphi_a + \frac{c^2}{g(\varphi_a)}\right\}$

This result must hold independently of the value of the external potential φ_a. This can only be true if the following important results hold:

$$D = c^2\nu$$

which thus justifies the setting of $c^2\nu$ as the trace of the matter stress-energy tensor; also

$$g(\varphi) = \frac{c^2}{\varphi'} \qquad (8.8)$$

where $$\varphi' = \varphi - \left(\varphi_a - \frac{c^2}{g(\varphi_a)}\right) \qquad (8.9)$$

Until now the gravitational potential has remained determined only up to an additive constant, as is customary. With the dependence of g on φ as given in equation (8.8), such an indeterminancy would lead to ambiguity in the force and field equations. Therefore one should look upon equation (8.9) as providing a unique scaling for the gravitational potential which is to be used in equation (8.8).

Combining these results, Nordström was able to recover the final field and force equations as

$$\Box^2\varphi' = -g(\varphi')\nu = -\frac{c^2}{\varphi'}\nu \qquad (8.1b)$$

$$R^g_\mu = g(\varphi')\nu\varphi'_{,\mu} = c^2\nu\,(\ln\varphi')_{,\mu} \qquad (8.2b)$$

where $c^2\nu$ is the trace of the matter stress-energy tensor. Because g is now a function of φ', the dependence of inertial rest mass m on gravitational potential must be recalculated and, in a straightforward manner, equation (8.3a) is replaced by the simpler

$$m \propto \varphi'_a \qquad (8.3b)$$

It should be noted that the modifications carried out here do more than just make the gravitational mass of a closed system proportional to its total energy. Rather they bring the theory into line with the full requirements of the democracy of energy-momentum in the sense of the previous chapter. In particular in the modified theory, all energy, including that of the gravitational field itself, contributes equally as a source of the gravitational field. This can be followed through in the derivation outlined above.

The energy of the gravitational field enters as the second term in the integral of equation (8.6) and thus, through equation (8.7), contributes to the gravitational mass of the entire system. Further, it can be seen that it is the presence of this second term which results in the dependence of g on φ in equation (8.8). Without this second term - that is if we ignored the contribution of gravitational field energy - we would conclude from a derivation similar to the one above that g was a constant independent of φ and we would be returned to a theory much like Nordström's first theory.

In other words, the fact that the total source term is $\frac{c^2\nu}{\varphi'}$, which is no longer directly proportional to ν, is an expression of the fact that the gravitational field associated with the source masses of ν is itself field producing.

Nordström concluded his paper by returning to Einstein's original objection. Because of the modification of his source term in line with Einstein's "Entwurf..." observations, the rotation of a spinning body or molecular motion within a body would no longer influence their acceleration in free fall. However, a simple calculation showed that the sideways velocity of a freely falling body would still influence its downward acceleration exactly in accord with equation (8.4) of the earlier theory, with the added requirement that g now vary as in equation (8.8).[13]

In Section 4.6.3, I stressed the importance of this last result as an indicator of Einstein's changing understanding of the fundamental empirical facts of gravitation. We can be sure that Einstein was aware that this result still held within Nordström's second theory. For in the last paragraph of his paper, Nordström noted that Einstein had shown that bodies with internal motions whose configurations are not stationary do not fall in accord with the requirements of the modified equation (8.4). In particular he had shown that the free fall trajectory of an elastically oscillating system varies with the phase of oscillation around an average value which is given by the modified equation (8.4).

8.4 The Significance of Stresses as Gravitational Sources

We have seen how Nordström was able to bring his theory into accord with the requirement of the equality of inertial and gravitational mass by adopting the source term suggested by Einstein. Further, and of greatest importance, Nordström incorporated a reply in his modified gravitation theory to Einstein's "Entwurf..." argument against the possibility of a consistent scalar gravitation theory.

---

13. See also G. Nordström, "Die Fallgesetze und Planetenbewegungen in der Relativitätstheorie", Annalen der Physik, 43 (1914), pp.1101-10.

To begin, Nordström presented an analysis of the gravitational interactions of a spherical electron, which he modelled as being held together by elastic stresses concentrated in its surface.[14] From this analysis Nordström was able to draw the conclusion that the radius of the electron must vary in inverse proportion with the gravitational potential, as scaled by equation (8.9). He then proceeded to a brief

---

14. Nordström, "Zur Theorie...", pp.540-3.

but lucid explanation of why this dependence of length on gravitational potential had to be regarded as a general property of all matter within his theory:

> "That the dependence of the length dimension of a body on the gravitational potential must be a general property of matter according to the theory developed here, has been proven by Hr. Einstein, while he has shown that otherwise an apparatus can be constructed with which one could pump energy out of the gravitational field. In Einstein's example one considers a non-deformable rod, which can be stretched moveably between two vertical rails. One could let the rod fall tensioned, then relax it and raise it again. Tensioned, the rod has a greater weight than untensioned, and therefore it would provide more work in falling than would be used in raising the untensioned rod. However, because of the lengthening of the rod in falling, the rails must be divergent and the excess work in falling will be used again through the work of the tensioning forces at the ends on the rod.
>
> Let S be the total tension (tension times cross-sectional area) of the rod and l its length. Because of the tension, the gravitating mass of the rod is increased by
>
> $$\frac{g(\Phi)}{c^2} Sl = \frac{1}{\Phi'} Sl$$
>
> In falling this gravitating mass provides the extra work
>
> $$-\frac{1}{\Phi'} Sl\, d\Phi'$$
>
> At the same time however the work
>
> $$S\,dl$$
>
> will be given up at the ends of the rod. The setting of both expressions as equal provides
>
> $$-\frac{1}{\Phi'} d\Phi' = \frac{1}{l} dl$$
>
> or integrated $l\Phi' =$ constant..."

To this he added a footnote:

> "If the rod is deformable, work will be expended in tensioning it and the rest energy of the rod correspondingly increased. Also through this the weight experiences an increase, which yields the extra work dA in falling. As, however, the rest energy decreases in falling, the work which was gained by relaxing the rod is smaller than that used in tensioning, and the difference amounts to precisely dA." [15]

15. "Dass nach der hier entwickelten Theorie die Abhängigkeit der Längendimensionen eines Körpers vom Gravitationspotential eine allgemeine Eigenschaft der Materie sein muss, hat Hr. Einstein bewiesen, indem er gezeigt hat, dass sich sonst eine Einrichtung konstruieren liesse, womit man aus dem Gravitationsfelde Energie auspumpen könnte. In dem Einsteinschen Beispiel betrachtet man einen nicht deformierbaren Stab, der zwischen zwei vertikalen Schienen beweglich eingespannt werden kann. Man könnte den Stab gespannt fallen lassen, ihn dann entspannen und wieder heben. Gespannt hat der Stab ein grösseres Gewicht als ungespannt, und er würde also beim Fallen eine grössere Arbeit leisten, als beim Heben des ungespannten Stabes verbraucht wird. Wegen der Verlängerung des Stabes beim Fallen müssen aber die Schienen divergent stehen, und der Überschuss an Arbeit beim Fallen wird durch die Arbeit der spannenden Kräfte an den Enden des Stabes wieder verbraucht.

Es sei S die Gesamtspannung (Spannung mal Querschnitt) des Stabes, l die Länge desselben. Wegen der Spannung ist die gravitierende Masse das Stabes um

$$\frac{g(\Phi)}{c^2} Sl = \frac{1}{\Phi'} Sl$$

vergrössert. Beim Fallen leistet diese gravitierende Masse die Mehrarbeit

$$-\frac{1}{\Phi'} Sl\, d\Phi' .$$

An den Enden des Stabes wird aber gleichzeitig die Arbeit

$$S dl$$

abgegeben. Das Gleichsetzen der beiden Ausdrücke leifert

$$-\frac{1}{\Phi'} d\Phi' = \frac{1}{l} dl$$

oder integriert $l\Phi' =$ konst..."

"Wenn der Stab deformierbar ist, wird beim Spannen desselben eine Arbeit aufgewandt, und die Ruhenergie des Stabes dementsprechend vergrössert. Auch hierdurch erfährt das Gewicht eine Zunahme, die beim Fallen eine Mehrarbeit dA ergibt. Da aber beim Fallen die Ruhenergie abnimmt, ist die Arbeit, die beim Entspannen des Stabes gewonnen wird, kleiner als die beim Spannen verbrauchte, und der Unterschied beträgt eben dA."

<u>Ibid</u>. pp.544-5.

This passage contains a statement of the thought experiment central to Einstein's "Entwurf..." objection, but reduced to its barest essentials. A rigid stressed rod provides more work in falling than is needed to raise it in its unstressed state, given that stresses do indeed have weight. The apparent contradiction can be resolved if it is assumed that rods lengthen with decreasing gravitational potential. If the rod is stressed, then energy will be needed to work against these stresses. To close the energy balance, we set this energy equal of the extra energy liberated by the extra weight of the stresses in the rod. As Nordström showed, this leads to a quite specific dependence of the length of rods on gravitational potential within his gravitation theory. His calculation follows directly from the basic equations and definitions of his modified theory and the result is

$$l \propto \frac{1}{\varphi}, \qquad (8.10)$$

where l is the length of the rod in question.

I should like to stress that the energy transfer described here seems to be more of a pencil and paper affair than a real process. For the energy provided by the lowering of the stresses in the rod never becomes available for any external use, but is regarded as absorbed immediately as the work required to lengthen the rod. This latter process, of course, never actually requires external work to be supplied either. If the rod is unstressed then no external work is needed. If the rod is stressed, the situation is the same; the energy required is thought of as coming from the lowering of the stresses themselves.

This in turn seems to weaken the earlier conclusion that "stresses have weight". Stresses do still have weight within the formalism of the theory. But this weight does not behave in the same way as that of normal masses. For lowering the weight of a normal mass in a gravitational field results in the liberation of energy which can be used in any selected process. This, as we have seen, is not the case with the weight of stresses, for lowering stresses yields no net usable energy.

This directly resolves Einstein's "Entwurf..." objection, at least as far as Nordström's theory is concerned. For it follows that no net work is required to lift the massless but stressed members of the mirrored containers described and, therefore, it is no longer possible to set up a cycle from which a net energy gain would follow.

Further to this, Nordström's account of this matter suggests that Einstein was the source of the hypothesis that the length of a rod varies with gravitational potential and the conclusion that this resolves the earlier objection, for he attributes the entire argument to Einstein in the passage quoted above. The argument is sufficiently different to that of the "Entwurf..." paper to make it unlikely that Nordström was referring to it. But nowhere in his publications does Einstein present the argument or even claim credit for it. In a number of places he retracts the original "Entwurf..." objection. But he does so with an unrevealing and matter of fact note that it can be resolved if the length of rods are allowed to vary with gravitational potential.[16]

Finally, Nordström continued his paper by showing that the dependence of the length of rods on gravitational potential was part of a wider set of dependences of physical quantities on gravitational potential. In particular, he demonstrated that time as measured by natural

---

16. See Einstein, "Entwurf...", p.261; A. Einstein, "Zum gegenwärtigen Stande des Gravitationsproblems", Physikalische Zeitschrift, 14 (1913), p.1253; A. Einstein, "Zum Relativitäts-problem", Scientia, 15 (1914), pp.343-4.

processes – the periods of clocks, t – vary with gravitational potential in exactly the same way:

$$t \propto \frac{1}{\varphi}, \qquad (8.11)$$

This followed directly from the constancy of the speed of light, a requirement which Nordström was most concerned to retain in his theory. For this speed could be measured by clocking the time a light pulse takes to traverse a measuring rod. If the length of the rod varies according to equation (8.10), then the period of the clock must vary according to equation (8.11), if the measured speed of light is to remain the same constant. Nordström also established this dependence by a direct analysis of the behaviour of two types of clocks, a gravitational clock consisting of orbiting bodies and a clock based on an oscillating mass.

The extensive dependence of quantities on the gravitational potential makes Nordström's theory, in this form, somewhat cumbersome conceptually. In the following year, Einstein was to find an alternative formulation of the theory which made the function of these dependences much more transparent. Before he arrived at this formulation, Einstein was to give another which, in its own right, was a remarkable condensation and simplification of Nordström's theory. This will be examined in the next section.

8.5 Einstein's 1913 Derivation of Nordström's Theory

By the time of his 1913 Vienna address, Einstein was satisfied that Nordström's modified theory answered all the objections which had been made against it. As we saw earlier (Section 6.1), his only objection was epistemological. He believed that the theory did not satisfy the requirements of Mach's critique of space; in particular it did not entail the relativity of inertia.

In this section I present an outline of Einstein's rederivation and reformulation of Nordström's theory as given in the 1913 Vienna address.[17] Einstein's development of the theory is far more compact than Nordström's and shows how the theory arises very naturally out of the attempt to construct a scalar gravitation theory within special relativity. In spite of this Nordström's derivation seems to be better known.[18] I hope that my account here will help to rectify this.

The starting point of the theory is the decision to represent the gravitational field by a single scalar quantity φ.* This suggests the fundamental action principle

$$\delta \int \varphi \, dt_r = 0 \tag{8.12a}$$

for the motion of a free point mass, or, equivalently,

$$\delta \int m_o \varphi \, dt_r = 0 \tag{8.12b}$$

where m_o is a mass constant peculiar to the body in question and independent of the gravitational potential. With the usual techniques this yields the force equation for the four-force F_μ

$$F_\mu = \frac{dP_\mu}{dt_r} = m_o c^2 \varphi_{,\mu} \tag{8.13}$$

---

17. Einstein, "Zum gegenwärtigen Stande...", pp.1251-4.

18. See for example Laue, "Die Nordströmsche..." op.cit.; Harvey, pp.454-5, 457-9.

* The comments on notation of Section 8.1 still hold here. In addition I use the subscript "r" to indicate that a quantity is evaluated in the rest frame and the subscript "o" to indicate that the quantity is proper or self-measured.

where the four-momentum of the body is given by

$$P_\mu = m_o \varphi U_\mu \tag{8.14}$$

and
$$U_\mu = \frac{dx_\mu}{dt_r}$$

This force equation is equivalent to Nordström's equation (8.2b). Even allowing for the fact that the latter deals with force densities, their equivalence is still masked by the dependence in Nordström's theory of rest mass m on gravitational potential according to equation (8.3b). If the dependence is made explicit as

$$m = m_o \varphi'_a \tag{8.3c}$$

and substituted into Nordstrom's equations, then the equivalence asserted above becomes manifest. This also applies to the remainder of the equations of Einstein's formulation.

At this point in the derivation, Einstein alluded to earlier work and the need to postulate that proper lengths and times vary with gravitational potential. He introduced the factor ω, an undetermined function of φ , which relates the rest coordinate measures of space and time increments with their proper or self-measured values:

$$dx_o{}^\mu = \omega \, dx_r{}^\mu$$

It must be noted that, within the logic of the derivation, this does not amount to the unsupported assumption that such an effect occurs. For the function is left undertermined and, as far as we are to know at this stage, could well be a constant, independent of φ .

Now it follows directly from the definition of the components of the stress-energy tensor that this tensor is given by

$$T_\mu^\nu = \rho_o \varphi \omega^3 U_\mu U^\nu \tag{8.15}$$

for the case of a frictionless fluid of density ρ. (Result 1. Appendix E) Focussing on this, the force equation (8.13) and (8.14) can be rewritten for this continuous case. Consider a volume element ΔV containing mass Δm_0. For this we have

$$\frac{1}{\Delta V_r}\frac{d}{dt_r}\left(\Delta m_0 \varphi U_\mu\right) = \frac{1}{\Delta V_r}\Delta m_0 c^2 \varphi_{,\mu}$$

which can be shown (Result 2. Appenix E) to be equivalent to

$$\left(\rho_0 \varphi \omega^3 U_\mu U^\nu\right)_{,\nu} = \rho_0 \omega^3 c^2 \varphi_{,\mu} \tag{8.16a}$$

Using the definition (8.15) we conclude

$$T^\nu_{\mu,\nu} = T^\nu_\nu \frac{1}{\varphi}\varphi_{,\mu} \tag{8.16b}$$

From this Einstein stepped directly to the field equation . Its form was specified by requiring them

(i) to be a natural generalisation of Poisson's equation

(ii) to use the trace T^ν_ν as a source term, on the basis of considerations similar to those in the "Entwurf..." paper

and (iii) to combine with equation (8.16b) to provide a law of conservation of total energy-momentum of the form

$$\left(T^\nu_\mu + t^\nu_\mu\right)_{,\nu} = 0 \tag{8.17}$$

where t^ν_μ is the stress-energy tensor of the gravitational field.

This immediately suggests the field equation

$$\varphi \Box^2 \varphi = \varphi \varphi^{,\nu}_{,\nu} = -\varkappa T^\nu_\nu \tag{8.1c}$$

where $\varkappa$ is a constant and $\varphi^{,\mu} = \eta^{\mu\nu}\varphi_{,\nu}$. ($\eta^{\mu\nu}$ is the metric tensor of Lorentzian spacetime.) The appropriateness of this choice can be confirmed by substituting this field equation into equation (8.16b) which yields

$$T^\nu_{\mu,\nu} + \frac{1}{\varkappa}\varphi^{,\nu}_{,\nu}\varphi_{,\mu} = 0$$

This can be rewritten in the form of equation (8.17) if we note that

$$\frac{1}{\varkappa}\varphi^{,\nu}_{,\nu}\varphi_{,\mu}=\frac{1}{\varkappa}\left[\varphi^{,\nu}\varphi_{,\mu}-\frac{1}{2}\delta^{\nu}_{\mu}\varphi^{,\tau}\varphi_{,\tau}\right]_{,\nu} \quad *$$

thus giving

$$t^{\nu}_{\mu}=\frac{1}{\varkappa}\left[\varphi^{,\nu}\varphi_{,\mu}-\frac{1}{2}\delta^{\nu}_{\mu}\varphi^{,\tau}\varphi_{,\tau}\right] \tag{8.18}$$

This concludes Einstein's derivation of Nordström's theory. It is the most compact demonstration which has yet come to hand of how Nordström's theory arises naturally and directly from the search for a scalar theory of gravitation within special relativity. However, the derivation is incomplete, for it is necessary for Einstein to introduce the exact functional dependence of ω on φ - that ω is directly proportional to φ - as an external result derived elsewhere. This in no way compromises the logical completeness of his derivation of the equations of the theory. A brief analysis of the arguments used shows that Einstein's derivations of all important equations - notably the conservation law (8.16b) and the field equation (8.1c) - are independent of the exact functional dependence of ω on φ. This in itself is an illuminating comment on the status of the dependence of lengths and times on gravitational potential within the theory.

What I wish to propose now is a way in which the exact functional dependence of ω on φ can be determined within the framework of Einstein's derivation. It will be recalled that a variation of length with gravitational potential only needed to be considered when stressed masses were considered in order to retain the conservation of energy within the theory. This suggests that the exact dependence of ω on φ can be determined by examining the conservation law for a stressed

---

* $\delta^{\nu}_{\mu} = 1$ if $\nu=\mu$

$\quad\; = 0$ if $\nu\neq\mu$

matter distribution.

Consider a frictionless fluid of mass density ρ under an isotropic pressure, whose self measured value is Π_o and which is constant at all points in the fluid. The stress-energy tensor for the fluid is composed of two terms.

$$T_{\mu}^{\nu} = M_{\mu}^{\nu} + S_{\mu}^{\nu} \tag{8.19}$$

where the matter term is

$$M_{\mu}^{\nu} = \rho_o \varphi \omega^3 U_{\mu} U^{\nu}$$

and the stress term is

$$S_{\mu}^{\nu} = \Pi_o \varphi \omega^3 \left[\frac{1}{c^2} U_{\mu} U^{\nu} - \delta_{\mu}^{\nu}\right]$$

which can be seen to be a natural generalisation of the usual special relativistic expression for this term.[19]

We have from equation (8.16a) that

$$M_{\mu,\nu}^{\nu} = M_{\nu}^{\nu} \frac{1}{\varphi} \varphi_{,\mu}$$

In Appendix E, Result 3 , I show that

$$S_{\mu,\nu}^{\nu} = S_{\nu}^{\nu} \frac{1}{\omega} \omega_{,\mu}$$

Finally we must insist that equation (8.16b) still holds for the combined stress-energy tensor given in equation (8.19). Indeed we would require it to hold for all stress-energy tensors, if energy-momentum conservation is to be maintained and the theory is to retain its character. Clearly this can only be the case if

$$\omega \propto \varphi \tag{8.20}$$

whereby we have recovered the length and time dependence of Nordström's theory as given in equations (8.10) and (8.11).

---

19. As in, for example, P.G. Bergmann, Introduction to the Theory of Relativity (New York: Dover, 1976), p.129.

8.6 The Curvature of Spacetime

So far Nordström's theory has been approached as a gravitation theory which is set in a spacetime whose structure is given by the special theory of relativity. However, the emergence of the gravitational potential dependence of the length of rods and periods of clocks compromises this view to a certain extent. For whilst we are still free to think of the spacetime in which Nordström's theory is set as Lorentzian, this structure is no longer directly accessible to us from the measurements of rods and clocks.

In 1914 Einstein and Fokker published a new account of Nordström's theory in which the significance of this feature was made especially transparent.[20] They relaxed the requirement that spacetime be Lorentzian and introduced a spacetime with the quadratic metric of the "Entwurf..." theory. They then specialised this spacetime with the requirement that it always be possible to choose a (global) reference system in which the speed of light has a constant value c. This led to the metric

$$ds^2 = \varphi^2 (c^2 dt^2 - dx^2 - dy^2 - dz^2) \tag{8.22}$$

where φ is an arbitrary scalar. Equating this scalar to the gravitational potential and assuming that the world lines of free masses are geodesics in this spacetime, they were able to show that the force equations of Nordström's theory immediately followed, using the techniques established in the "Entwurf..." paper. Moreover, the gravitational potential dependence of the length of rods and periods of clocks followed immediately from the above metric.

---

20. A. Einstein and A.D. Fokker, "Die Nordströmsche Gravitationstheorie vom Standpunkt des absoluten Differentialkalküls", Annalen der Physik, 44 (1914), pp.321-8.

But the question of a field equation still remained. Carried forward by their success so far in finding a generally covariant formulation of Nordström's theory, Einstein and Fokker asked what generally covariant equations were possible candidates for the field equation. They were uniqely drawn to

$$R = \varkappa T$$

where R is the Riemann curvature scalar, T the trace of the stress-energy tensor and $\varkappa$ a constant. They immediately confirmed that this was equivalent to the field equation of Nordström's theory.

For Einstein the most significant aspect of this account was the power of the absolute differential calculus to generate a viable gravitation theory from the principle of the constancy of the velocity of light, with virtually no further assumptions. In particular, he seemed impressed by the role of the Riemann curvature scalar in the field equation. This clearly reflected back onto his work on developing the "Entwurf..." theory and the construction of field equations within it.

8.7 Conclusion

We can look upon this last formulation of Nordström's theory as the final vindication of Einstein's 1907 belief that no scalar theory of gravitation was possible within special relativity. We have seen how the sustained attempt to construct such a theory inexorably led Nordström to climb out of the spacetime structure of special relativity into a more general one.

Further, this result came solely from the search for a (special) relativistically acceptable gravitation theory. In the course of the chapter we have seen how constraints virtually identical to those of the constraint enevelope of the last chapter circumscribed and directed the development of Nordström's theory. Its original form came directly from the search for a special relativistic gravitation theory with the correct Newtonian limit. The introduction of the trace of the stress-energy tensor as the source term resuluted directly from the requirement of the democracy of energy-momentum. For only then could all the energy of a body, and not just its rest energy, be a source of the gravitational field. Nordström's derivation of his second theory clearly showed how gravitational field energy itself now came to contribute equally with all other forms of energy as a source of the gravitational field - exactly as required by the democracy of energy-momentum. Finally the introduction of the crucial effect, the dependence of the length of rods on gravitational potential, was precipitated by the requirement of the conservation of energy. For if stresses were to have weight, as implied by the new source term, then such an effect would be necessary to prevent a breakdown of this conservation law.

Finally we discover that, out of this process, we arrive at a theory whose structure is remarkably like that of Einstein's final general theory of relativity. Both are generally covariant theories in which gravitation is dealt with as the curvature of a spacetime, whose structure is governed by field equations based on the Riemann curvature tensor. Just about the only difference between them is that one is scalar theory and the other a second rank tensor theory.

Thus we find both Einstein's and Nordström's theory starting out at the same point: the search for a relativistically acceptable gravitation theory. Einstein's programme became dominated rapidly first by the belief that no such theory is possible within special relativity and then by the grand vision of a theory which would realise the general relativity of motion and satisfy the requirements of Mach's epistemological critique of space. Nordström, however, continued to work only within the limitations of the original goal - that of constructing a relativistically acceptable gravitation theory, preferably within the confines of special relativity. Then we find both converging to remarkably similar theories. The difference between them seems only to be that Einstein was induced to step up to a second rank tensor theory, whilst Nordström was content to remain on the simpler level of a scalar theory. With such an outcome, it is hard to escape the conclusion that the requirement that each theory contain a relativistically acceptable gravitation theory has decisively conditioned the form of the final theories.

CHAPTER 9

CONCLUSION

9. CONCLUSION

Einstein's passage from his first speculations on the problem of gravitation in relativity theory in 1907 to his final general theory of relativity in 1915 is one of the most remarkable episodes in the history of science. What is striking about it is Einstein's persistent ability to anticipate and insist upon the most far-reaching of results well before they had become apparent to his colleagues. Already in 1907 he was convinced that gravitation could not be contained within the theory of space and time of special relativity. At the same time he recognised that the key to the phenomenon of gravitation lay in the interpretation of the full significance of the equality of gravitational and inertial mass and the related law of Galileo that all bodies fell with equal accelerations. This led him to seek a theory which incorporated the gravitational field into the structure of space and time - and later spacetime - itself. Finally, Einstein's belief in the need for an extension of the theory of space and time of special relativity was underpinned throughout by his conviction that the theory contained a fundamental defect, the prior and preferred status of inertial frames of reference. Out of this Einstein sought to construct a theory which would display the physical equivalence of all frames of reference. This led him to seek a generally covariant theory and to turn to the mathematical tools which were peculiarly suited to the task, the tensor calculus of Ricci and Levi-Civita.

However, this process was surrounded by a wider network of concepts and arguments, many of which were neither complete nor successful. Einstein's original 1907 argument for the inability of special relativity to contain an acceptable gravitation theory was based on the assumption that such a theory must imply that the downward acceleration of a falling body is independent of its sideways velocity. Einstein's final general theory of relativity does not entail such a result. However, on first analysis, it does seem possible that a second rank tensor theory still within special relativity may well entail it. Later, Einstein's conclusion was vindicated by more sustained analyses. Nordström, for example, working with Einstein's critical assistance, developed the possibilities of a scalar theory of gravitation within special relativity at the same time as Einstein was developing his general theory of relativity. The theory which resulted clearly showed how gravitation must burst out of special relativity.

Further, in 1907, Einstein took the significance of the equality of inertial and gravitational mass to lie in a physical equivalence or indistinguishability of the effects of acceleration and gravitation. In his work on gravitation prior to 1913, he hoped to use this as a bridge over which he could pass from the accessible structure of inertial fields to the more obscure structure of gravitational fields. However, his final general theory of relativity strongly limits the prospects of such a device. Inertial fields produced by acceleration are capable only of mimicking the gross features of gravitational fields, but do not mirror their finer structure. Moreover Einstein invoked the equivalence as holding between uniform acceleration and homogeneous gravitational fields. Within the final theory such "homogeneous gravitational fields" would not even be regarded to be gravitational fields at all. This led Einstein to conclude that static gravitational fields were spatially Euclidean, in contradiction to his final theory, and was directly responsible for his recovery of the incomplete "half deflection" angle for the degree of bending of starlight by the sun's gravitational field. Finally Einstein carried this conclusion over to the early versions of his general theory of relativity in 1913, where it contributed to the notorious rejection of the Ricci tensor as the basis of the field equations.

Einstein's search for a generally covariant theory of space, time and gravitation began as a part of a wider epistemologically based critique of existing theories. The critique required the final elimination of space as a causally active agent in physical theory. It involved the establishment of the general relativity of the motion of bodies and a new account of the origin of the inertial forces which act upon them as arising from a gravitational interaction with all the other masses of the universe.

However, Einstein's critique was both ill-born and ill-fated. It had arisen through a misunderstanding of the import of Mach's critique of the Newtonian concept of space and was carried through in a corpuscular world view which saw the only causally active entities in the universe to be particles in motion in a passive and featureless container space. In particular, this view was essential to the recovery of his account of the origin of inertia from his critique. The overthrow of this world view was only then being consummated by his own theory of relativity. This theory promoted the view that space had to be seen as permeated with causally active fields. Indeed, Einstein even came to concede a view quite antithetical to his original critique: that the general theory of relativity in promoting the concept of the unified field, led to the view that space may well be the fundamental medium of all reality.

Whilst such results were emerging, Einstein engaged in a subtle but persistent programme of reassessment and modification of the basic concepts and arguments which had initially guided him. By 1918 he no longer stated the principle of equivalence as asserting the indistinguishability of acceleration and gravitation. Rather he saw it as asserting the essential identity of inertial and gravitational mass, an identity which found expression in the theory in the fact that gravitational phenomena, as well as inertial phenomena, were accounted for by the metric tensor.

Einstein's original demand for the final elimination of space as a causal entity gave way to a weaker but related requirement - that space no longer be a cause that acts but is not acted upon itself. This found satisfactory expression in the "no prior geometry" feature of his final theory. The demand for the general relativity of motion and the physical equivalence of all frames of reference gave way to the requirement of the formal equivalence of all coordinate systems for the writing of the laws of nature and the insistence on generally covariant formulation of physical theories. Einstein had to abandon his early ideas on the origin of inertia, in spite of intensive attempts to transfer them to the new field theoretic framework. He conceded their incongruity in this new framework, although he never doubted the essential soundness of the ideas which gave birth to them.

Thus we can see that the heuristics which guided and directed Einstein in his discovery of the general theory of relativity formed an organic network of ideas and concepts, which was continuously growing and evolving, even to the point where they contradicted their original forms. In spite of this, I have sought to argue that we can still understand the emergence of the theory as the rational outcome of a single and well defined process, provided we recognise that, beneath Einstein's grand vision of a theory of the general relativity of motion, lay a more mundane objective, that of constructing a relativistically acceptable gravitation theory. Indeed I have argued that, provided the requirements of this objective are carried through to the full, one is led inexorably to the final general theory of relativity or one very much like it.

Thus I have sought to characterise the principle of equivalence, the concept of the general relativity of motion and the many related considerations which surrounded them as a force which drove Einstein forward in the search for his final theory and which suggested concepts and devices crucial to the development of this theory. This process was carried out within the constraint envelope provided by the requirement that the final theory contain a relativistically acceptable gravitation theory. Einstein refused to compromise this constraint enevelope, even when he thought it required him to give up his cherished ideal, a theory in which all the basic laws, including those governing the gravitational field are generally covariant. The rigidity of this framework acted as a counter to the volatility of the driving heuristic and provided a stable framework within which the theory could develop. It guaranteed that if Einstein was ever to reach a satisfactory conclusion to his work, then it would be with his final general theory of relativity or one very much like it.

Finally, on the simpler but parallel level of a scalar theory of gravitation, the case of the development of Nordström's theory of gravitation illustrates the power of the constraint enevelope to lead to a theory of the nature of Einstein's final general theory of relativity. Nordström sought to construct a scalar theory of gravitation within special relativity. With Einstein's critical assistance, the sustained application of consistency requirements similar to those in Einstein's constraint enevelope led to a final theory which had burst out of the bounds of special relativity. In its most transparent form, the theory accounted for gravitation as the curvature of spacetime, as is the case in the general theory of relativity, and with a field equation based on the Riemann curvature tensor.

BIBLIOGRAPHY

BIBLIOGRAPHY

This bibliography is not an exhaustive listing of all the journal articles and monographs cited in the text. Rather it consists of a selection of primary and secondary sources which comprise the main field in which the historical investigations of the text were carried out. For a critical survey of the general resources available to the historian interested in the general theory of relativity, see L. Pyenson, "The Goettingen Reception of Einstein's General Theory of Relativity", diss. John Hopkins University 1974. Also a convenient microprint collection of all Einstein's published and some other works is available as The Collected Writings of Albert Einstein (New York: Readex Microprint Corporation, 1960).

PRIMARY SOURCES

Abraham, M. "Zur Theorie der Gravitation". Physikalische Zeitschrift,13(1912), pp.1-4, 176.

Abraham, M. "Das Elementargesetz der Gravitation". Physikalische Zeitschrift,13(1912), pp.4-5.

Abraham, M. "Der Freie Fall". Physikalische Zeitschrift,13(1912), pp.310-11.

Abraham, M. "Die Erhaltung der Energie und der Materie in Schwerkraft Felde". Physikalische Zeitschrift,13(1912), pp.311-4.

Abraham, M. "Das Gravitationsfeld". Physikalische Zeitschrift,13 (1912), pp.793-7.

Abraham, M. "Nochmals Gravitation und Relativität. Bemerkungen zu A. Einsteins Erwiderung". Annalen der Physik,39(1912), pp.444-8.

Abraham, M. "Relativität und Gravitation.Erwiderung auf eine Bemerkung des Hrn. A. Einstein". Annalen der Physik,38(1912), pp.1056-8.

Abraham, M. "Die Neue Mechanik". Rivista di Scienza,15(1914), pp.8-27

Abraham, M. "Neuere Gravitationstheorien". Jahrbuch der Radioaktivität und Elektronik,11(1914), pp.470-520.

Behacker, M. "Der freie Fall und die Planetenbewegung in Nordströms Gravitationstheorie". Physikalische Zeitschrift,14(1913), pp.989-92.

Born, M. The Born-Einstein Letters. Trans. I. Born, New York: Walker, 1971.

Einstein, A. "Zur Elektrodynamik bewegter Körper". Annalen der Physik,17(1905), pp.891-921.

"On the Electrodynamics of Moving Bodies" in Sommerfeld, ed., The Principle of Relativity, pp.37-65.

Einstein, A. "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen". Jahrbuch der Radioaktivität und Elektronik,4(1907), pp.411-62; 5(1908), pp.98-9.

Einstein, A. "Zum Ehrenfestschen Paradoxon". Physikalische Zeitschrift,12(1911), pp.509-10.

Einstein, A. "Über den Einfluss der Schwerkraft auf die Ausbreitung des Lichtes". Annalen der Physik,35(1911), pp.898-908.

"On the Influence of Gravitation on the Propagation of Light" in Sommerfeld, ed., The Principle of Relativity, pp.99-108.

Einstein, A. "Lichtgeschwindigkeit und Statik des Gravitationsfeldes". Annalen der Physik,38(1912), pp.355-69.

Einstein, A. "Zur Theorie des Statischen Gravitationsfeldes". Annalen der Physik,38(1912), pp.443-58.

Einstein, A. "Relativität und Gravitation: Erwiderung auf eine Bemerkung von M. Abraham". Annalen der Physik,38(1912), pp.1059-64.

Einstein, A. "Bemerkung zu Abraham's vorangehender Auseinandersetzung 'Nochmals Relativität und Gravitation'". Annalen der Physik,39(1912), p.704.

Einstein, A. "Gibt es eine Gravitationswirkung, die der electrodynamischen Induktionswirkung analog ist?" Vierteljahrsschrift für gerichtliche Medizin,44(1912), pp.37-40.

Einstein, A. "Physikalische Grundlagen einer Gravitationstheorie". Naturforschende Gesellschaft Vierteljahrsschrift (Zurich), 58(1913), pp.284-90.

Einstein, A. "Die formale Grundlage der allgemeinen Relativitätstheorie". Preussische Akademie der Wissenschaften, Sitzungsberichte,1914 Part 2, pp.1030-85.

Einstein, A. "Zum Relativitäts-problem". Scientia,15(1914), pp.337-348.

Einstein, A. "Nachtragliche Antwort an Herrn Reissner". Physikalische Zeitschrift,15(1914), pp.108-110.

Einstein, A. "Prinzipielles zur verallgemeinerten Relativitätstheorie und Gravitationstheorie". Physkalische Zeitschrift,15(1914), pp.176-80.

Einstein, A. "Zum gegenwärtigen Stande des Gravitationsproblems". Physikalische Zeitschrift,14(1913), pp.1249-66.

Einstein, A. "Zur allgemeinen Relativitätstheorie". Preussische Akademie der Wissenschaften, Sitzungsberichte, 1915 Part 2, pp.778-86, 799-81.

Einstein, A. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie". Preussische Akademie der Wissenschaften, Sitzungsberichte,1915 Part 2, pp.831-9.

Einstein, A. "Die Feldgleichungen der Gravitation". Preussische Akademie der Wissenschaften, Sitzungsberichte, 1915 Part 2, pp.844-7.

Einstein, A. "Die Grundlage der allgemeinen Relativitätstheorie". Annalen der Physik,49(1916), pp.769-822.

"The Foundation of the General Theory of Relativity" in Sommerfeld, ed., The Principle of Relativity, pp.109-64.

Einstein, A. "Über Friedrich Kottlers Abhandlung 'Über Einsteins Äquivalenzhypothese und die Gravitation'". Annalen der Physik, 51(1916), pp.639-42.

Einstein, A. "Ernest Mach". Physikalische Zeitschrift,17(1916), pp.101-4.

Einstein, A. "Naherungsweise Integration der Feldgleichungen der Gravitation". Preussische Akademie der Wissenschaften, Sitzungsberichte, 1916 Part 1, pp.688-96.

Einstein, A. "Hamiltonsches Prinzip und allgemeine Relativitätstheorie". Preussische Akademie der Wissenschaften, Sitzungsberichte, 1916 Part 2, pp.1111-6.

"Hamilton's Principle and the General Theory of Relativity" in Sommerfeld, ed., The Principle of Relativity, pp.165-73.

Einstein, A. Über die spezielle und die allgemeine Relativitätstheorie: Gemeinverständlich. Braunschweig: Vieweg, 1917.

Relativity: The Special and the General Theory: A Popular Exposition. Trans. by R.W. Lawson, 1920; 15th enl. ed. rpt. London: Methuen, 1977.

Einstein, A. "Kosmologische Betrachtungen zur allgemeinen Relativätstheorie". Preussische Akademie der Wissenschaften, Sitzungsberichte, 1917 Part 1, pp.142-52.

"Cosmological Considerations on the General Theory of Relativity". in Sommerfeld, ed., The Principle of Relativity, pp.175-88.

Einstein, A. "Prinzipielles zur allgemeinen Relativitätstheorie". Annalen der Physik,55(1918), pp.240-4.

Einstein, A. "Dialog über Einwände gegen die Relativitätstheorie". Die Naturwissenschaften,6(1918), pp.697-702.

"Dialogue on Objections to the Theory of Relativity". Trans. by H.M. Schwartz in G.E. Tauber, ed., Albert Einstein's Theory of Relativity,New York: Crown, 1979, pp.99-108.

Einstein, A. "Kritisches zu einer von Herrn de Sitter gegebenen Lösung der Gravitationsgleichungen". Preussische Akademie der Wissenschaften, Sitzungsberichte,1918 Part 2, pp.270-2.

Einstein, A. "Fundamental Ideas and Methods of Relativity Theory, Presented in their Development". Unpublished manuscript in German at Einstein Archives, Princeton, circa 1919. Translation of title and selected passage in G. Holton,"Finding Favour with the Angel of the Lord. Notes towards the Physchobiographical Study of Scientific Genius" in Y. Eklana, ed., The Interaction between Science and Philosophy, Humanities Press, 1975, pp.369-71.

Einstein, A. "Spielen Gravitationsfelder im Aufbau der materiellen Elementarteilchen eine wesentliche Rolle?" Preussische Akademie der Wissenschaften, Sitzungsberichte, 1919 Part 1, pp.349-56.

"Do Gravitational Fields Play an Essential Part in the Structure of the Elementary Particles of Matter?" in Sommerfeld, ed., The Principle of Relativity, pp.189-98.

Einstein, A. Äther und Relativitätstheorie. Berlin: Springer, 1920.

"Ether and Relativity". Trans. by G.B. Jeffrey and W. Perrett in Sidelights on Relativity, London: Methuen, 1922.

Einstein, A. "Meine Antwort auf die antirelativitätstheoretische G.m.b.H". Berliner Tageblatt und Handelszeitung, 27 August, 1920, pp.1-2.

Einstein, A. "Inwiefern lässt sich die moderne Gravitationstheorie ohne die Relativität begründen?". Die Naturwissenschaften, 8 (1920), pp.1010-1.

"To what Extent can the Modern Theory of Gravitation be Formulated without Relativity". Trans. in G.E. Tauber, ed., Albert Einstein's Theory of General Relativity, New York: Crown, 1979, pp.94-7.

Einstein, A. The Meaning of Relativity, 1922; 6th. rev. ed., Chapman and Hall: London, 1976.

Einstein, A. "Geometrie und Erfahrung". Preussische Akademie der Wissenschaften, Sitzungsberichte, 1921 Part 1, pp.123-30.

"Geometry and Experience". Trans. by G.B. Jeffrey and W. Perrett in Sidelights on Relativity, London: Methuen, 1922.

Einstein, A. "On the Theory of Relativity". 1921, in Ideas and Opinions, London: Souvenir, 1973, pp.246-9.

Einstein, A. "Grundgedanken und Probleme der Relativitätstheorie" in Nobelstiftelsen, Les Prix Nobel en 1921-1922. Stockholm: Imprimerie Royale, 1923.

Einstein, A. "Fundamental Ideas and Problems of the Theory of Relativity", in G.E. Tauber, ed., Albert Einstein's Theory of General Relativity, New York: Crown, 1979, pp.51-8.

Einstein, A. "Über den Äther". Schweizerische Naturforschende Gesellschaft, Verhandlungen, 105 (1924), pp.85-93.

Einstein, A. "Newtons Mechanic und ihr Einfluss auf die Gestaltung der Theoretischen Physik". Naturwissenschaften, 15 (1927), pp.273-6.

"The Mechanics of Newton and their Influence on the Development of Theoretical Physics". 1927, in Ideas and Opinions. London: Souvenir, 1973, pp.253-61.

Einstein, A. "Space, Ether and the Field in Physics". Forum Philosophicum, 1 (1930), pp.180-4.

Einstein, A. "Maxwell's Influence on the Evolution of the Idea of Physical Reality". 1931, in Ideas and Opinions, London: Souvenir, 1973, pp.266-70.

Einstein, A. "Notes on the Origin of the General Theory of Relativity". 1933, in Ideas and Opinions, London: Souvenir, 1973, pp.285-90.

Einstein, A. Review of R.C. Tolman: Relativity, Thermodynamics and Cosmology. Science, 80 (1935), p.358.

Einstein, A. "Physics and Reality". 1936, in Ideas and Opinions, London: Souvenir, 1973, pp.290-323.

Einstein, A. and Infeld, L. The Evolution of Physics. Cambridge: Cambridge Univ. Press, 1938.

Einstein, A. "The Fundaments of Theoretical Physics". 1940, in Ideas and Opinions, London: Souvenir, 1973, pp.323-35.

Einstein, A. Autobiographical Notes. 1949; rpt. La Salle & Chicago, Illinois: Open Court, 1979.

Einstein, A. "On the Generalized Theory of Gravitation". 1950, in Ideas and Opinions, London: Souvenir, 1973, pp.341-56.

Einstein, A. Foreword to M. Jammer, Concepts of Space, 1953; rpt. Cambridge, Mass: Harvard Univ. Press, 1969.

Einstein, A. Ideas and Opinions. 1954; rpt. London: Sourvenir/Condor, 1973.

Einstein, A. and Grossmann, M. "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation". Zeitschrift für Mathematik und Physik, 62 (1913), pp.225-261.

Einstein, A. and Fokker, A.D. "Die Nordströmsche Gravitationstheorie vom Standpunkt des absoluten Differentialkalküls". Annalen der Physik, 44 (1914), pp.321-8.

Einstein, A. and Grossman, M. "Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie", Zeitschrift für Mathematik und Physik, 63 (1914), pp.215-25.

Kottler, F. "Uber Einsteins Äquivalenzhypothese und die Gravitation". Annalen der Physik, 50 (1916), pp.955-72.

Kretschmann, E. "Über den physikalischen Sinn der Relativitätspostulate A. Einsteins neue und seine ursprüngliche Relativitätstheorie". Annalen der Physik,53(1917), pp.575-614.

Laue, M. "Zur Dynamik der Relativitätstheorie". Annalen der Physik, 35(1911), pp.524-42.

Laue, M. Die Relativitätstheorie. Vol I & II, Braunschweig: Vieweg, 1952.

Laue, M. "Die Nordströmsche Gravitationstheorie". Jahrbuch der Radioaktivität und Elektronik,14(1917), pp.263-313.

Lorentz, H.A. "Considerations on Gravitation". Proceedings of the Royal Academy of Sciences, Amsterdam.2(1900), pp.559-574.

Lorentz, H.A. "Alte und neue Fragen der Physik". Physikalische Zeitschrift,11(1910), pp.1234-57.

Mach, E. History and Root of the Principle of the Conservation of of Energy. 1872 with notes 1909, trans. P. Jourdain, Chicago: Open Court, 1911.

Mach, E. The Science of Mechanics: A Critical and Historical Account of its Development. 1883; trans. by T. J. McCormack, 6th ed., 9th German ed., La Salle, Illinois: Open Court, 1974.

Mach, E. The Analysis of Sensations. 1886; 6th ed., 1902; trans. C.M. Williams and S. Waterlow, New York: Dover, 1959.

Mach, E. "The Economical Nature of Physical Inquiry". 1896, in J.J. Kockelmans (ed.) The Philosophy of Science, New York: Free Press, 1968, pp.174-87.

Mie, G. "Bemerkungen zur Einsteinschen Gravitationstheorie". Physikalische Zeitschrift,15(1914), pp.115-22, 169-76, 263.

Minkowski, H. "Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Köpern." Nachrichten von der königlichen Gesellschaft der Wissenschaft zu Göttingen, Mathematisch-Physikalische Klasse (1908), pp.53-111.

Minkowski, H. "Raum und Zeit". Physikalische Zeitschrift,10(1909) p.104. "Space and Time" in Sommerfeld, ed., The Principle of Relativity, pp.75-91

Newcombe, S. The Elements of the Four Inner Planets and the Fundamental Constants of Astronomy. Supplement to the American Ephemeris and Nautical Almanac for 1897. Washington: Government Printing Office, 1895.

Nordström, G. "Relativitätsprinzip und Gravitation". Physikalische Zeitschrift,13(1912), pp.1126-9.

Nordström, G. "Träge und Schwere Masse in der Relativitätsmechanik". Annalen der Physik,40(1912), pp.856-78.

Nordström, G. "Zur Theorie der Gravitation vom Standpunkt des Relativitätsprinzips". Annalen der Physik,42(1913), pp.533-54.

Nordström, G. "Über den Energiesatz in der Gravitationstheorie". Physikalische Zeitschrift,15(1914), pp.375-80.

Nordström, G. "Die Fallgesetze und Planetenbewegungen in der Relativitätstheorie". Annalen der Physik,43(1914), pp.1101-10.

Planck, M. "Zur Dynamik bewegter Systeme". Annalen der Physik,26 (1908), pp.1-34.

Poincaré, H. "Sur la Dynamique d l'Électron". Comptes-Rendus hebdomadaires des séances de l'Academie des Science. Paris, 140(1905), pp.1504-8 (short version); Rendiconti del Circolo Matematico di Palermo,21(1906), pp.129-75 (complete). Sec. 1-5, 9 trans. to English in C.W. Kilmister, Special Theory of Relativity. Oxford: Pergamon, 1970, pp.144-85.

de Sitter, W. "On Einstein's Theory of Gravitation and its Astronomical Contents". Monthly Notices of the Royal Astronomical Society,76(Suppl.1916), pp.699-728; 77(1916), pp.155-84; 78(1917), pp.3-28.

de Sitter, W. "On the Relativity of Rotation in Einstein's Theory". Proceedings of the Academy of Science, Amsterdam,19(1917), pp.527-32.

de Sitter, W. "On the Relativity of Inertia. Remarks on Einstein's Latest Hypothesis". Proceedings of the Academy of Science, Amsterdam,19(1917), pp.1217-25.

de Sitter, W. "On the Curvature of Space". Proceedings of the Academy of Science, Amsterdam,20(1917), pp.229-42.

de Sitter, W. "Further Remarks on the Solutions of the Field Equations of Einstein's Theory of Gravitation". Proceedings of the Academy of Science, Amsterdam,20(1918), pp.1309-12.

Sommerfeld, A. "Zur Relativitätstheorie I & II". Annalen der Physik,32(1910), pp.749-76; 33(1910), pp.649-89.

Sommerfield, A. (ed.) The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity by H.A. Lorentz, A. Einstein, H. Minkowski and H. Weyl. Trans. by W. Perrett and G.B. Jeffrey, 1923; rpt. New York: Dover, 1952.

Speziali, P. (ed.) Albert Einstein-Michele Besso Correspondance 1903-1955, Paris: Herman, 1972.

Speziali, P. "Einstein Writes to his Best Friend" in A.P. French,ed. Einstein. A Centenary Survey. Cambridge, Mass: Harvard Univ. Press, 1979, pp.263-9.

SECONDARY SOURCES

Bergmann, O. "Scalar Field Theory as a Theory of Gravitation". American Journal of Physics,24 (1956), p.38.

Berkson, W. Fields of Force. London: Routledge & Kegan Paul, 1974.

Berry, A. A Short History of Astronomy. 1898; rpt. New York: Dover, 1961.

Bunge, M. "Mach's Critique of Newtonian Mechanics". American Journal of Physics,34(1966), pp.585-96.

Chazy, J. Théorie de la Relativité et la Mecanique Céleste. Paris: Gauthier - Villars, 1928.

Clark, R.W. Einstein: The Life and Times. London: Hodder and Stoughton, 1979.

Dorling, J. "Did Einstein need General Relativity to solve the Problem of Absolute Space? Or had the Problem already been solved by special Relativity?". British Journal for the Philosophy of Science,29(1978), pp.311-23.

Earman, J. and Glymour, C. "Einstein and Hilbert: Two Months in the History of General Relativity". Archive for the History of the Exact Sciences,19(1978), pp.291-308.

Earman, J. and Glymour, C. "Lost in the Tensors: Einstein's Struggles with Covariance Principles 1912-1916". Studies in the History and Philosophy of Science,9(1978), pp.251-78.

Earman, J. and Glymour, C. "Relativity and Eclipses: The British Eclipse Expedition of 1919 and their Predecessors". Historical Studies in the Physical Sciences,11(1980), pp.49-85.

Earman, J. and Glymour, C. "The Gravitational Red Shift as a Test of General Relativity: History and Analysis". Studies in the History and Philosophy of Science,11(1980), pp.175-214.

Frank, P. Einstein: His Life and Times. London: Jonathan Cape, 1948.

Fock, V. The Theory of Space, Time and Gravitation. London: Pergamon, 1959.

Gardner, M.R. "Relationism and Relativity". British Journal for the Philosophy of Science,28(1977), pp.215-33.

Goenner, H. "Mach's Principle and Einstein's Theory of Gravitation". Boston Studies in the Philosophy of Science,6(1970), pp.200-15.

Golden,N.J. Some Aspects of Mach's Principle within the Theory of General Relativity. Diss. Univ. of Wyoming 1971.

Graves, J.C. The Conceptual Foundations of Contemporary Relativity Theory. Cambridge, Mass.: M.I.T. Press, 1971

Harvey, A.L. "Brief Review of Lorentz - Covariant Scalar Theories of Gravitation". American Journal of Physics, 33, (1965), pp.449-60.

Herman, A. (ed.) Albert Einstein Arnold Sommerfeld Briefwechsel. Basel: Schwabe, 1968.

Hesse, M. Forces and Fields. Totowa, N.J.: Littlefield, Adams & Co., 1965.

Hirosige, T. "The Ether Problem, the Mechanistic World View and the Origins of the Theory of Relativity". Historical Studies in the Physical Sciences, 7 (1976), pp.3-82.

Hoffmann, B. "Einstein and Tensors". Tensor, 26 (1972), pp.157-62.

Hoffmann, B. Albert Einstein. Paladin, 1975.

Holton, G. "Mach, Einstein and the Search for Reality". Daedalus, 97 (1968), pp.636-73.

Illy, J. "The Birth of Einstein's Theory of Relativity". Manuscript, Budapest, 1976. (I am indebted to Mr. Illy for sending me a copy of this manuscript.)

Jammer, M. Concepts of Force. Cambridge, Mass: Harvard U.P., 1957.

Jammer, M. Concepts of Mass. Cambridge, Mass: Harvard U.P., 1961.

Jammer, M. Concepts of Space. 2nd ed., Cambridge, Mass: Harvard U.P., 1969.

Kahn, C. and Kahn, F. "Letters from Einstein to de Sitter on the Nature of the Universe". Nature, 257 (1975), pp.451-4.

Klein, O. "On the Foundations of General Relativity Theory and the Cosmological Problem" in G.E. Tauber, ed., Albert Einstein's Theory of General Relativity. New York: Crown, 1979.

Lanczos, C. "Einstein's Path from Special to General Relativity" in L.O'Raifeartaigh, ed., General Relativity: Papers in Honour of J.L. Synge, Oxford: Clarendon, 1972, pp.5-19.

Lanczos, C. The Einstein Decade (1905-1915). London: Elek Science, 1974.

Mehra, J. Einstein, Hilbert, and the Theory of Gravitation. Dordrecht: Riedel, 1974.

McCormmach, R. "Editor's Foreward". Historical Studies in the Physical Sciences, 7 (1976), pp, (xi)-(xxxv).

North, J.D. The Measure of the Universe. Oxford: Clarendon, 1965.

Norton, J. "Einstein, Mach's Principle and the Origins of the General Theory of Relativity". Proceedings of the Australasian Association for the History, Philosophy and Social Studies of Science, 12 (1980-1) Forthcoming.

Oppenheim, S. "Kritik des Newtonschen Gravitationsgesetzes" in Encyklopädie der Mathematischen Wissenschaften mit Einschluss ihrer Anwendungen. Leipzig: B.G. Teubner, 1922-34 (concluded in Feb.1920), pp.80-158.

Pauli, W. Theory of Relativity. 1921, annotated 1956; trans. G. Field, Oxford: Pergamon, 1958.

Pirani, F.A.E. "On the Definition of Inertial Systems in General Relativity". Helvetica Physica Acta, 1956 Suppl.IV, pp.198-203.

Pyenson, L. "The Goettingen Reception of Einstein's General Theory of Relativity". Diss. John Hopkins University 1974.

Pyenson, L. "Hermann Minkowski and Einstein's Special Theory of Relativity". Archive for History of Exact Sciences, 17 (1977), pp.71-95.

Reinhardt. M. "Mach's Principle - A Critical Review." Zeitschr. Naturforsch., 28 (1973), pp.529-37.

Roxburgh, I.W. and Tavakol, R. "The Gravitation Theories of Poincaré and Milne and the Non-Riemannian Kinematic Models of the Universe". Mon. Not. R. Astron. Soc., 170 (1975), pp.599-610.

Ritz, W. "Die Gravitation", Rivista di Scienza, 5 (1909), pp.241-255.

Sachs, M. The Field Concept in Contemporary Science. Springfield, Ill.: Charles C. Thomas, 1973.

Sachs, M. "On the Logical Status of Equivalence Principles in General Relativity Theory ". British Journal for the Philosophy of Science, 27 (1976), pp.225-9.

Schwartz, H.M. "Poincaré's Rendiconti Paper on Relativity". American Journal of Physics, 39 (1971), pp.1287-94; 40 (1972), pp.862-72, 1282-7.

Schwartz, H.M. "Einstein's Comprehensive 1907 Essay on Relativity". American Journal of Physics, 45 (1977), pp.512-7, 811-7, 899-902.

Seelig, C. Albert Einstein. A Documentary Biography. Transl. M. Savill, London: Staples 1956, p.76.

Sewell, W.C. "Einstein, Mach and the General Theory of Relativity". Diss. Case Western Univ. 1975.

Sklar, L. Space, Time and Spacetime. Berkeley: University of California Press, 1977.

Stachel, J. "Einstein's Odyssey. His Journey from Special to General Relativity". The Sciences (March 1979), pp.14-15, 32-34.

Stachel, J. "The Genesis of General Relativity". Lecture Notes in Physics 100 Einstein Symposium, Berlin, Berlin: Springer-Verlag, 1979.

Stachel, J. "Einstein and the Rigidly Rotating Disk" in A. Held, ed., General Relativity and Gravitation. A Hundred Years After the Brith of Einstein Vol.I New York: Plenum, 1979(?).

Stachel, J. "Einstein's Search for General Covariance, 1912-1915". Talk given at the Ninth International Conference on General Relativity and Gravitation, Jena, DDR. July 17, 1980. (Draft Manuscript)

Strauss, M. "Einstein's Theories and the Critics of Newton". Synthese, 18 (1968), pp.251-84.

Stein, H. "Some Philosophical Prehistory of General Relativity". Minnesota Studies in the Philosophy of Science, 8 (1977), pp.3-49.

Teske, A. "Einstein and Smoluchowski". Sudhoffs Archiv, 53 (1969), pp.292-305.

Thirring, W.E. "An Alternative Approach to the Theory of Gravitation". Annals of Physics, 16 (1961), pp.96-117.

Tonnelat, M. Histoire du Principe de Relativité. Paris: Flammarion, 1971.

Vizgin, V.P. and Smorodinskii Ya. A. "From the Equivalence Principle to the Equations of Gravitation". Soviet Physics Uspekhi, 22 (1979), pp.489-513.

Wheeler, J.A. "Einstein's Last Lecture" in G.E. Tauber, ed., Albert Einstein's Theory of General Relativity, New York: Crown, 1979, pp.187-90.

Whitrow, G.J. and Morduch, G.E. "General Relativity and Lorentz Invariant Theories of Gravitation". Nature, 188 (1960), pp.790-4.

Whitrow, G.J. and Morduch G.E. "Relativiistic Theories of Gravitation, a Comparative Analysis with Particular Reference to Astronomical Tests". Vistas in Astronomy, 6 (1965), pp.1-67.

Whitrow, G.J. "From the Problem of Fall to the Problem of Collapse: Three Hundred Years of Gravitation Theory." Philosophical Journal, 14 (1977), pp.67-84.

Whittaker, E.T. A History of Theories of Aether and Electricity: The Modern Theories 1900-1926. London: Nelson, 1953.

Zahar, E. "Mach, Einstein and the Rise of Modern Science". British Journal for the Philosophy of Science, 28 (1977), pp.195-213.

Zahar, E. "Einstein, Meyerson and the Role of Mathematics in Physical Discovery". British Journal for the Philosophy of Science, 31 (1980), pp.1-43.

Zenneck, J. "Gravitation" in Encyklopädie der Mathematischen Wissenschaften mit Einschluss ihrer Anwendungen. Leipzig: B.G. Teubner, 1903-21 (concluded in Aug.1901), pp.25-67.

APPENDICES

Appendix A Scalar, Vector and 2-Tensor Force Laws

The entries in Table 3.1 are drawn from the following considerations. Consider a body of rest mass m with instantaneous velocity v in the y direction alone in a static gravitational field which acts only in the x-direction. F_μ is the four-force acting on the body and $f_1 = -\sqrt{1-v^2/c^2}\, F_1$ is the x-component of the three dimensional force. (For simplicity in the 2-tensor case, the field is assumed to be symmetric in the y-direction.) For this case, from the expression for four-force, $F_\mu = m\dfrac{d^2x_\mu}{d\tau^2}$, where τ is proper time, it follows that

$$F_i = -\frac{m}{1-v^2/c^2}\frac{d^2x^i}{dt^2} \qquad \text{(A.1)}$$

where $i = 1,2,3$

Scalar Theory General force law: $F_\mu = m\,G_\mu$ (3.19a)

Therefore, for this case, $F_1 = mG_1$

and $f_1 = -m\sqrt{1-v^2/c^2}\, G_1$

Using (A.1), the x-acceleration is $\dfrac{d^2x}{dt^2} = -\left(1-v^2/c^2\right) G_1$

Vectory Theory General force law: $F_\mu = m\,U^\alpha G_{\alpha\mu}$ (3.19b)

For this case $U_1 = U_3 = 0$

Therefore $F_1 = m\left[U^0 G_{01} + U^2 G_{21}\right]$

But $F_\mu U^\mu \equiv 0$. Therefore $U^\alpha U^\mu G_{\alpha\mu} \equiv 0$. Therefore $G_{\alpha\mu} = -G_{\mu\alpha}$

Since the field acts only in the x-direction

$F_2 = m\left[U^0 G_{02} + U^1 G_{12}\right] = 0$

for a body with x-velocity only and for all $\mathcal{U}^1$.

Therefore $G_{12} = 0$. Therefore $G_{21} = 0$.

Hence $F_1 = m\mathcal{U}^0 G_{01}$

and $f_1 = -m G_{01}$ since $\mathcal{U}^0 = \frac{1}{\sqrt{1-V^2/c^2}}$

From (A.1) $\frac{d^2x}{dt^2} = -\sqrt{1-V^2/c^2}\, G_{01}$

2-Tensor Theory General force law $F_\mu = m\,\mathcal{U}^\alpha \mathcal{U}^\beta G_{\alpha\beta\mu}$ (3.19c)

For this case, $\mathcal{U}_1 = \mathcal{U}_3 = 0$

Therefore $F_1 = m\left[\mathcal{U}^0\mathcal{U}^0 G_{001} + \mathcal{U}^0\mathcal{U}^2(G_{021} + G_{201}) + \mathcal{U}^2\mathcal{U}^2 G_{221}\right]$

I now show that all but the first term in this expression is zero.

$F_\mu \mathcal{U}^\mu \equiv 0$. Therefore $\mathcal{U}^\alpha \mathcal{U}^\beta \mathcal{U}^\mu G_{\alpha\beta\mu} \equiv 0$

Therefore $(G_{\alpha\beta\mu} + G_{\beta\alpha\mu} + G_{\mu\alpha\beta} + G_{\alpha\mu\beta} + G_{\beta\mu\alpha} + G_{\mu\beta\alpha})$
$= 0$ (A.2)

In this case, a body with arbitrary motion has

$$F_2 = m\mathcal{U}^\alpha \mathcal{U}^\beta G_{\alpha\beta 2}$$
$$= m\left[\ldots + \mathcal{U}^1\mathcal{U}^2(G_{122} + G_{212}) + \mathcal{U}^0\mathcal{U}^1(G_{012} + G_{102}) + \ldots\right] = 0$$

Therefore $G_{122} + G_{212} = 0$, which gives, with (A.2), $G_{221} = 0$

Also $G_{012} + G_{102} = 0$

In this case, a body with arbitrary motion has

$$F_0 = m_0 \mathcal{U}^\alpha \mathcal{U}^\beta G_{\alpha\beta 0} = m\left[\ldots + \mathcal{U}^1\mathcal{U}^2(G_{120} + G_{210}) + \ldots\right]$$

From the assumption of symmetry in the y-direction, it follows that F_0 can have no terms <u>linear</u> in $\mathcal{U}^2$.

Therefore $G_{120} + G_{210} = 0$

Now, from (A.2)

$$G_{021} + G_{201} = -(G_{012} + G_{102} + G_{210} + G_{120})$$
$$= 0, \quad \text{using the above results.}$$

Combining we recover for the case in question

$$F_1 = m\,U^0 U^0\, G_{001}$$

and thus $f_1 = \dfrac{-m}{\sqrt{1 - v^2/c^2}}\, G_{001}$

From (A.1) $\dfrac{d^2x}{dt^2} = -G_{001}$

Appendix B Geodesics of a Weak, Static Gravitational Field

Consider a weak static gravitational field acting in the x-direction alone. We write its metric as

$$ds^2 = A\,dt^2 - B\,dx^2 - D\,dy^2 - E\,dz^2 \qquad \text{(B.1)}$$

where A,B,D and E are functions of x only.

The geodesics are given by

$$\frac{d}{d\tau}\frac{\partial L}{\partial \dot{x}^\mu} = \frac{\partial L}{\partial x^\mu} \quad \text{where} \quad L = A\dot{t}^2 - B\dot{x}^2 - D\dot{y}^2 - E\dot{z}^2$$

$$\dot{t} = \frac{dt}{d\tau} \quad \text{etc.}$$

and proper time τ satisfies $d\tau = \frac{ds}{c}$

Solving for $\mu = 0$ and $\mu = 1$ and combining, we recover the relation

$$\frac{dV_x}{dt} = -\frac{1}{2B}\frac{\partial A}{\partial x} + \left[\frac{1}{A}\frac{\partial A}{\partial x} - \frac{1}{2B}\frac{\partial B}{\partial x}\right]V_x^2 + \frac{1}{2B}\frac{\partial D}{\partial x}V_y^2 + \frac{1}{2B}\frac{\partial E}{\partial x}V_z^2$$

where V_x, V_y and V_z are the x, y & z components of the velocity of a freely falling test body. For the weak field metric (4.6a), $\frac{\partial D}{\partial x} = \frac{\partial E}{\partial x} \neq 0$ and thus the x-acceleration varies with sideways velocity V_y and V_z. This is not the case for the metric (4.6b), for this metric yields

$$\frac{\partial D}{\partial x} = \frac{\partial E}{\partial x} = 0 \,.$$

Appendix C Einstein's 1912 Scalar Theory of Gravitation

I examine the behaviour of spacetime and the matter contained within it for the special case of a spacetime governed by the metric

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 \qquad \text{(C.1)}$$

where c = c(x,y,z) is a function of the space coordinates x,y and z only. I show the results correspond to the predictions of Einstein's 1912 scalar theory of static gravitational fields, as developed in his "Lichtgeschwindigkeit und Statik des Gravitationsfeldes", Annalen der Physik, 38 (1912), pp.355-69 and "Zur Theorie des statischen Gravitationsfeldes", Annalen der Physik, 38 (1912), pp.443-58 (cited here as I and II respectively).

Behaviour of Rods, Clocks and Light Pulses

Equation (C.1) implies that

(i) the lengths of measuring rods are unaffected by the presence of the field and the spatial geometry determined by them is Euclidean

(ii) the period of a resting clock dT differs from coordinate time according to $dT = \frac{c}{c_L} dt$

where c_L is the speed of light in gravitation free space, and

(iii) a light pulse, which propagates along a null geodesic, will have a speed c, as measured by coordinate time and distance, and the speed will be isotropic. It can be confirmed from reference I that Einstein's theory makes identical predictions.

Particle Kinematics

All non-zero Christoffel symbols of the second kind

$$\left\{ {\alpha \atop \mu\nu} \right\} = \frac{1}{2} g^{\alpha\beta} \left(g_{\mu\beta,\nu} + g_{\beta\nu,\mu} - g_{\mu\nu,\beta} \right)$$

are

$$\left.\begin{aligned} \left\{ {0 \atop 0i} \right\} = \left\{ {0 \atop i0} \right\} &= \frac{1}{c} \frac{\partial c}{\partial x^i} \\ \text{and} \qquad \left\{ {i \atop 00} \right\} &= c \frac{\partial c}{\partial x^i} \end{aligned}\right\} \qquad \text{(C.2)}$$

$$i = 1, 2, 3$$

Substitution into the equation for the geodesic

$$\frac{d^2 x^\mu}{ds^2} + \left\{ {\mu \atop \alpha\beta} \right\} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds} = 0$$

yields

In x^i: $$\frac{d}{dt}\left(\frac{\dot{x}^i/c}{\sqrt{1 - V^2/c^2}} \right) = \frac{-\frac{\partial c}{\partial x^i}}{\sqrt{1 - V^2/c^2}} \qquad \text{(C.3)}$$

where $\dot{x}^i = \frac{dx^i}{dt}$ and $i = 1, 2, 3$

In t: $$\frac{d}{dt}\left(\frac{c}{\sqrt{1 - V^2/c^2}} \right) = 0 \qquad \text{(C.4)}$$

(C.3) corresponds exactly to Einstein's equation of motion (I.6a) and (C.4) to his relation (I.7).

Dynamic Quantities

The four-momentum $p^{\mu} = m\frac{dx^{\mu}}{d\tau}$ of a body which has rest mass m in a region remote from the gravitational field is given by

$$P_0 = (\text{energy}) = c_L \frac{mc}{\sqrt{1 - v^2/c^2}} \qquad \text{(C.5)}$$

$$-P_i = (i\text{-momentum}) = c_L \frac{\dot{x}^i/c}{\sqrt{1 - v^2/c^2}} \qquad \text{(C.6)}$$

$$(i = 1, 2, 3)$$

Allowing that Einstein sets $c_L = 1$,

(C.5) corresponds to Einstein's expression for the energy of a particle (I.8) and suggests the general result of (I. § 3) that the energy of a system be given by c times its local or self measured energy.

(C.6) agrees with Einstein's choice of the term to represent momentum in his (I.6b).

Curvature Tensor and its Contractions

The Riemann curvature tensor is given by

$$R^{\alpha}{}_{\beta\gamma\delta} = \left\{ {\alpha \atop \beta\delta} \right\}_{,\gamma} - \left\{ {\alpha \atop \beta\gamma} \right\}_{,\delta} + \left\{ {\tau \atop \beta\delta} \right\} \left\{ {\alpha \atop \tau\gamma} \right\} - \left\{ {\tau \atop \beta\gamma} \right\} \left\{ {\alpha \atop \tau\delta} \right\}$$

Substituting values for the Christoffel symbols from equations (C.2) the following are recovered as the only non-zero terms

$$\left.\begin{aligned} R^{i}{}_{0j0} &= -R^{i}{}_{00j} = c\,\frac{\partial^2 c}{\partial x^i \partial x^j} \\ R^{0}{}_{ij0} &= -R^{0}{}_{i0j} = \frac{1}{c}\,\frac{\partial^2 c}{\partial x^i \partial x^j} \\ & i,j = 1,2,3 \end{aligned}\right] \qquad \text{(C.7)}$$

The Ricci tensor is given by

$$R_{\mu\nu} = R^{\tau}{}_{\mu\tau\nu}$$

and from (C.7) it follows that its only non-zero terms are

$$\left.\begin{aligned} R_{00} &= c\nabla^2 c \\ R_{ij} &= -\frac{1}{c}\,\frac{\partial^2 c}{\partial x^i \partial x^j} \\ & i,j = 1,2,3 \end{aligned}\right] \qquad \text{(C.8)}$$

From this it follows that the curvature scalar $R = R^{\tau}_{\tau}$ is given by

$$R = \frac{2}{c}\nabla^2 c \qquad \text{(C.9)}$$

Field Equations

For the case of a source mass free gravitational field, Einstein's field equations, as given in references I and II, do not yield the same results as the field equations of the final general theory of relativity. In this latter theory, for this case,

$$R_{\mu\nu} = 0$$

From equations (C.7) and (C.8), this condition can be seen to be equivalent to

$$R^{\alpha}{}_{\beta\gamma\delta} = 0$$

Thus, for the case of a source mass free field, the general theory of relativity predicts that the field can only occur in a flat spacetime and, solving these equations, we find that c can vary linearly at most with the spatial coordinates x^i, where i=1,2,3.

It is interesting to note that the field equation of reference I, which had to be rejected because of a dynamical inconsistency, corresponds to the equation

$$R = \varkappa T$$

where T is the trace of the stress-energy tensor and $\varkappa$ a constant.

Appendix D   Deflection of Light in a Schwarzschild Field

Based on an extension of the argument presented by J. Weber in H.Y. Chiu and W.F. Hoffmann (eds.), Gravitation and Relativity (New York: Benjamin, 1964), pp.231-3

In isotropic coordinates the Schwarzschild field of a mass M, located at the origin of coordinates (x,y,z,t), is approximately given by

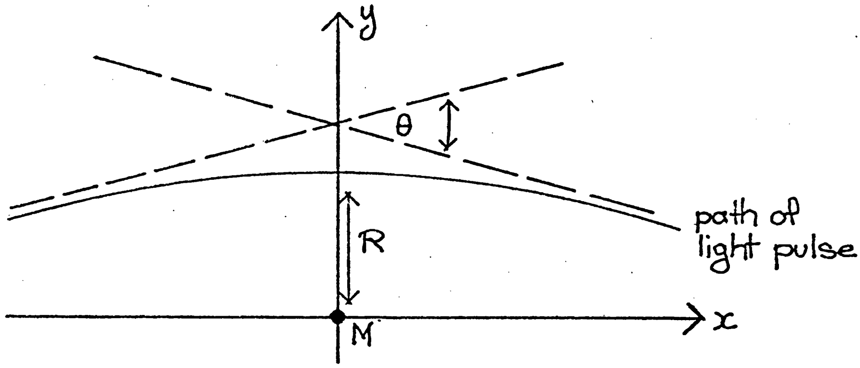
$$ds^2 = c^2\left(1 - A\frac{2GM}{c^2 r}\right)dt^2 - \left(1 + B\frac{2GM}{c^2 r}\right)\left(dx^2 + dy^2 + dz^2\right) \qquad \text{(D.1)}$$

to first order in $\frac{2GM}{c^2 r}$, where G = gravitational constant
$r^2 = x^2 + y^2$
c = speed of light in gravitation free space

and A = B = 1

The factors A and B have been introduced to enable determination of the contribution of each term to the final result.

Consider a light pulse moving roughly parallel to the x-axis in the xy plane, approaching to within R of the centre of the mass at the origin and being deflected by an angle θ



The y-component of the geodesic of the light pulse is

$$\frac{d^2y}{dk^2} + \left\{ {y \atop \mu\nu} \right\} \frac{dx^\mu}{dk}\frac{dx^\nu}{dk} = 0 \qquad \text{(D.2)}$$

where k is a path parameter

and $\left\{ {\alpha \atop \mu\nu} \right\} = \frac{1}{2} g^{\alpha\beta}\left(g_{\mu\beta,\nu} + g_{\beta\nu,\mu} - g_{\mu\nu,\beta}\right)$

Taking k = ct as the path parameter, we have $\frac{dt}{dk} = \frac{1}{c}$

Assuming that the deflection is small, we have $\frac{dx}{dk} = 1$ and we

can ignore any terms containing the velocity $\frac{dy}{dt}$ as small.

Thus (D.2) becomes

$$\frac{d^2y}{dx^2} = -\frac{1}{c^2}\left\{ {y \atop tt} \right\} - \left\{ {y \atop xx} \right\}$$

which reduces to

$$\frac{d^2y}{dx^2} = (A+B)\,\frac{GMy}{c^2(x^2+y^2)^{3/2}}$$

after evaluation and substitution of the Christoffel symbols.

By assuming that y differs little from R, this can be integrated directly to yield the deflection angle θ:

$$\Delta\theta = \left(\frac{dy}{dx}\right)_{x=\infty} - \left(\frac{dy}{dx}\right)_{x=-\infty} = \int_{-\infty}^{\infty} \frac{d^2y}{dx^2}\,dx$$

$$= (A+B)\,\frac{GMR}{c^2}\int_{-\infty}^{\infty} \frac{dx}{(x^2+R^2)^{3/2}} = (A+B)\,\frac{2GM}{c^2R} \qquad \text{(D.3)}$$

Thus the total deflection, for A = B = 1, is $4\frac{GM}{c^2R}$ and it can be seen that the time and space coefficients of the metric each contribute an equal half deflection $2\frac{GM}{c^2R}$

Appendix E Notes to Einstein's Derivation of Nordström's Theory

1. Stress-Energy Tensor for frictionless mass distribution of density ρ.

T_{μ}^{ν} = Flux of P_{μ} momentum in x^{ν} direction

$$= \frac{dP_{\mu}}{dx^1 dx^2 dx^3} \frac{dx^{\nu}}{dx^0} = \frac{dP_{\mu}}{dV_r} \frac{dx^{\nu}}{dt_r} = \rho_r \varphi \, U_{\mu} U^{\nu} = \rho \varphi \omega^3 U_{\mu} U^{\nu}$$

2. Transition of force equation to continuous case of frictionless fluid

First we have $\frac{d}{dt_r}\left(\frac{1}{\Delta V_r}\right) = -\frac{1}{\Delta V_r^2} \frac{d}{dt_r}\left(\frac{\Delta t \Delta V}{\Delta t_r}\right)$ since $\Delta t \Delta V = \Delta t_r \Delta V_r$

$$= -\frac{1}{\Delta V_r^2} \frac{\Delta t \Delta V}{\Delta V_r} \frac{d}{dt_r} \frac{\Delta x^{\nu}}{\Delta t_r}$$

$$= -\frac{1}{\Delta V_r} \frac{dt_r}{dx^{\nu}} \frac{d}{dt_r}\left(\frac{dx^{\nu}}{dt_r}\right) \text{ when } \Delta x^{\nu} \text{ is very small}$$

$$= -\frac{\omega^3}{\Delta V_0} \left(U^{\nu}_{,\nu}\right)$$

Therefore $\frac{\Delta m_0}{\Delta V_r} c^2 \varphi_{,\mu} = \frac{1}{\Delta V_r} \frac{d}{dt_r} (\Delta m_0 \varphi \, U_{\mu})$

becomes

$$\rho_0 \omega^3 c^2 \varphi_{,\mu} = (\rho_0 \varphi \omega^3 U_{\mu} U^{\nu})_{,\nu} - \rho_0 \omega^3 \varphi \, U_{\mu} (U^{\nu}_{,\nu}) - \Delta m_0 \varphi \, U_{\mu} \frac{d}{dt_r}\left(\frac{1}{\Delta V_r}\right)$$

$$= (\rho_0 \varphi \omega^3 U_{\mu} U^{\nu})_{,\nu}$$

3. Stress-energy tensor for a frictionless fluid under constant pressure

Preliminary results:

(i) Since the pressure is constant, the world line of individual particles in the fluid will be given by equations (8.13) and (8.14)

i.e. $\frac{d}{dt_r}(\varphi\, U_\mu) = c^2 \varphi_{,\mu}$

or $\varphi\, U_{\mu,\nu} U^\nu + U_\mu \varphi_{,\nu} U^\nu = c^2 \varphi_{,\mu}$

(ii) Consider a small rigid box of self measured volume ΔV_0 in motion. By definition $\frac{d\Delta V_0}{dt_r} = 0$

Now $U^\nu(\omega^3)_{,\nu} = \frac{d}{dt_r}\omega^3 = \Delta V_0 \frac{d}{dt_r}\frac{1}{\Delta V_r} = -\omega^3 (U^\nu_{,\nu})$

using the calculation of Result 2

Finally $S^\nu_{\mu,\nu} = \left(\Pi_0 \varphi \omega^3 \left[\frac{1}{c^2} U_\mu U^\nu - \delta^\nu_\mu\right]\right)_{,\nu}$

$$= \Pi_0 \Big(\frac{\omega^3}{c^2} U_\mu \varphi_{,\nu} U^\nu + \varphi (\omega^3)_{,\nu} \frac{1}{c^2} U_\mu U^\nu + \frac{\omega^3}{c^2} U_{\mu,\nu} \varphi U^\nu + \frac{\omega^3}{c^2} \varphi U_\mu U^\nu_{,\nu} - (\omega^3 \varphi)_{,\mu} \Big)$$

$= -3\Pi_0 \omega^3 \varphi \frac{1}{\omega} \omega_{,\mu}$ after reduction with (i) and (ii)

$$= S^\nu_\nu \frac{1}{\omega} \omega_{,\mu}$$