Title
Unification and explanation in early Kaluza-Klein theories

Author
Muntean, Ioan Lucian

Publication Date
2009

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Unification and Explanation in Early Kaluza-Klein Theories

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Philosophy

by

Ioan Lucian Muntean

Committee in charge:

Craig Callender, Chair
William Bechtel
Nancy Cartwright
Ken Intriligator
Christopher Smeenk
Christian Wüthrich

2009

The dissertation of Ioan Lucian Muntean is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2009

TABLE OF CONTENTS

SIGNATURE PAGE iii
TABLE OF CONTENTS iv
LIST OF ABBREVIATIONS vii
LIST OF SYMBOLS viii
LIST OF FIGURES ix
LIST OF TABLES x
ACKNOWLEDGEMENTS xi
VITA xvi
ABSTRACT OF THE DISSERTATION xvii

PART I. PHILOSOPHICAL PERSPECTIVES ON UNIFICATION 1

CHAPTER 1. ARGUMENTS FOR UNIFICATION 2

1.1. Some possible definitions of unification 4
1.2. Unity of a theory and unificatory theory 9
1.3. Two attitudes towards unification 15
1.4. Strong philosophical positions about unity and unification 18
1.5. From unification to realism: M. Friedman (1983) 24
1.6. The quest after unification in science 34

CHAPTER 2. ARGUMENTS AGAINST UNIFICATION 42

2.1. Philosophical skepticism against unification 42
2.2. Scientists against unified theories and unification 57
2.3. The ideal of "modest unificationism" and a rebuttal 65

CHAPTER 3. UNIFICATION AND EXPLANATION: M. FRIEDMAN AND P. KITCHER 69

3.1. The D-N account of explanation 69
3.2. Explanation and counting phenomena: Friedman (1974) 73
3.3. Patterns of argumentation: Kitcher (1976, 1981) 79

CHAPTER 4. CAUSAL EXPLANATION VERSUS UNIFICATORY EXPLANATION (KITCHER, WOODWARD) 84

4.1. Unificationism: explanation without causation (Kitcher, 1989) 84
4.2. Woodward's three criticisms against unificationism 91
4.3. Two replies to Woodward: diversity in explanation 96
4.4. Interventions, modularity and unification 102
4.5. Pluralism about explanation and causation 105

CHAPTER 5. A DEFLATIONARY ACCOUNT OF UNIFICATION (M. MORRISON) 108

5.1. Two types of unity 113
5.2. Is causation the missing "machinery"? 119
5.3. Morrison's argument revisited 122

CHAPTER 6. WHY A PHILOSOPHICAL APPRAISAL OF KALUZA AND KLEIN?....... 126

6.1. Two approaches to unification 126
6.2. The place of my case study within existing approaches....... 131
6.3. The relevance of Kaluza-Klein unification 133
Summary of Part I 139

PART II. KALUZA AND KLEIN THEORIES....... 142

CHAPTER 7. UNIFIED THEORIES AT THE BEGINNING OF THE 20TH CENTURY....... 144

7.1. The covariant *EM*: an unificatory program 145
7.2. EM as a synthetic unification (N. Maxwell and M. Morrison) 161

CHAPTER 8. HOW TO MESH TOGETHER *EM* AND *GR*?....... 165

8.1. Differences between GR and EM....... 169
8.2. Similarities between GR and EM 174
8.3. Gravity and electromagnetism on a different par: Einstein Field Equations....... 177
8.4. A uniform description of EM and GR: the Einstein-Hilbert action 180
8.5. Weyl's unification and his non-Riemannian metric....... 188

CHAPTER 9. GEOMETRIZATION OF PHYSICAL FORCES 194

9.1. Geometrization of gravitation in GR....... 202
9.2. Geometrization of EM in GR 206

CHAPTER 10. KALUZA'S FIVE-DIMENSIONAL THEORY....... 211

10.1. Geometrization in Kaluza and its historical background 216
10.2. Logical structure of Kaluza's argument 224
10.3. Details of Kaluza's assumptions 225
10.4. Kaluza's results 230
10.5. Problems with Kaluza's theory: covariation of x^4 and geodesics....... 233

CHAPTER 11. KLEIN AND QUANTUM MECHANICS IN FIVE DIMENSIONS (1926)....... 237

11.1. Historical background 237
11.2. Klein's metric 246
11.3. Klein's second argument: compactification and waves in 5-D....... 252
11.4. The new argument with COMP....... 256

PART III. THE KALUZA-KLEIN UNIFICATION AND EXPLANATION....... 260

CHAPTER 12. UNIFICATION OF PHYSICAL INTERACTIONS 264

12.1. Economy and duality of representations 264
12.2. Unification and physical interactions 268
12.3. Inter-theoretical relations....... 273
12.4. Reductionism: electromagnetism and gravitation 278
12.5. The reduction of EM to GR....... 289

12.6. T. Maudlin's ranking of unifications ... 292

CHAPTER 13. INGREDIENTS OF KALUZA'S UNIFICATION ... 299

13.1. The Campbell-Magaard theorem ... 302
13.2. Step 1: The "make room" procedure ... 306
13.3. Step 2: Identifications of Christoffel symbols ... 308
13.4. Does Kaluza's theory achieve unification? ... 319

CHAPTER 14. UNIFICATION IN KLEIN ... 325

14.1. Invariance and factorization of the 5-D action ... 328
14.2. The wavefunction in 5-D as an external element of unification ... 332
14.3. Symmetries in Klein's theory ... 334
14.4. Symmetries and ground states: Kaluza versus Klein ... 340
14.5. Klein's novel type of unification and Kitcher revisited ... 345

CHAPTER 15. EXPLANATION, SPACETIME AND DIMENSIONALITY ... 351

15.1. Formalism as *explanans* in spacetime theories ... 353
15.2. Is Minkowski spacetime properly an *explanans?* ... 357
15.3. Non-causal explanations and spacetime ... 361
15.4. Interpretations of the theoretical structure in Kaluza and Klein ... 364
15.5. Are false theories explanatory? ... 368
15.6. An excurse in "Hyperspace" ... 371
15.7. Dimensional explanation ... 382
15.8. "Dimensionally challenged" models (dimensional truncation) ... 388
15.9. A "bare model" of the 5-D spacetime ... 394

CHAPTER 16. DIMENSIONAL EXPLANATION IN KALUZA-KLEIN THEORIES ... 399

16.1. Kaluza's dimensional explanation ... 403
16.2. The heuristics of Klein's compactification ... 404
16.3. What does Klein's compactification of x^4 explain? ... 406
16.4. The explanation of charge quantization: an unexpected *explanandum* ... 408
16.5. Quantum mechanics and the Kaluza-Klein unification ... 413
16.6. Particles as *explananda* in Klein ... 419
16.7. Symmetries of EM as *explananda* in Klein ... 425
16.8. The stability of Klein's vacuum ... 427
16.9. Valediction ... 432

APPENDIX. DIFFERENTIAL GEOMETRY: METRIC AND MANIFOLD ... 435

The manifold ... 435
Riemannian spaces ... 440
Transformation of metric from one system to the other ... 442

REFERENCES ... 444

LIST OF ABBREVIATIONS

CM = Classical Mechanics, in its Newtonian formulation or in its Hamilton-Jacobi-Lagrange formulation

D=2 = a theory that assumes that spacetime has two space dimensions

D=3 = a theory that assumes that spacetime has three space dimensions

D-N = the deductive-nomological model of explanation as formulated by C. Hempel

EM = Theory of Electromagnetism, especially in its covariant formulation

GR = Theory of General Relativity

QM = A generic term referring to generic quantum mechanics (no interpretations are assumed here)

PDE = "Partial Differential Equations" or systems of such equations

ODE = "Ordinary Differential Equations" or systems of such equations

TOE = Theory of Everything

UFT = Unified Field Theory(ies)

4-D, 4D = a theory that presupposes a four dimensional manifold with three spatial dimensions and on time dimension

5-D, 5D = Kaluza's or Klein's theory that presuppose a five dimensional manifold with four space dimensions and one time dimension

5D+ = Any theory that assumes that the spacetime manifold has more than five spacetime dimensions

LIST OF SYMBOLS

$\mathbf{x}^4$ = the fifth direction of the 5-D manifold.

$\mathbf{x}^0$ = the temporal axis of the 4-D and 5-D manifolds associated to time t.

LIST OF FIGURES

Figure 1 (elevated rectangles represent quantum theories)...................................... 37

Figure 2 Types of interactions among theories.. 273

LIST OF TABLES

Table 1: Types of geometries in Weyl and Cartan 190

Table 2: The dimensional explanation in Kaluza and Klein............................... 403

ACKNOWLEDGEMENTS

I owe the greatest debt to Craig Callender for guiding me along the sinuous path of assembling this dissertation and for being the driving force behind this project. I will always be grateful for his help in adapting to the high demands of UCSD right from the beginning of my graduate years. Craig made me believe in the relevance of the philosophy of science and philosophy of physics to philosophy writ large. Craig was enthusiastic about my project and encouraged me to write a dissertation in philosophy of physics even when I was not convinced to do so. Craig carefully read drafts of papers and parts of the thesis even if at first appeared desultory and rambling and provided invaluable guidance on how to improve and polish philosophical arguments. Since I have started working with Craig, my writing has improved due to his diligent counseling. Without his thoughtful guidance advice and constructive criticism this dissertation would be left waning.

Much gratitude is due to Nancy Cartwright, from whom I have learned much about how to write and build philosophical arguments. Nancy welcomed me as a student when I did not have a clue what path to follow in the philosophy of science. She nurtured my interest in the study of laws of nature in scientific thinking and from here I moved toward the idea of scientific unification. Nancy illustrates academic citizenship and friendship. She built a strong community of philosophers of science at UCSD with a strong emphasis on the philosophy of social sciences: I wish I would have more time to learn about this fascinating topic. Nancy is an example how important is to be passionate about doing philosophy of science and believe in its impact on the society in general.

Special thanks go to William Bechtel who initiated me in the philosophy of mechanisms and other exciting parts of philosophy of science. I owe a lot to his expertise in the philosophy of psychology, cognitive sciences as well to mentoring me in teaching philosophy of science at any level. I believe that my interest in philosophy of biology is due to his talent in teaching this fascinating subject.

Christian Wüthrich helped me with the form and content of my dissertation by carefully reading several of its incarnations. I am grateful that he accepted joining my committee in mid course and am inspired by his expertise in General Relativity and Quantum Gravity and by his elegant writing style as well. He rightly pointed out flaws in the argument and helped me in preparing materials for publication. His constructive criticism has been inestimable in preparing the final form of the dissertation.

I owe deep gratitude to the faculty and colleagues in the Department of Philosophy at UCSD who made me feel at home for almost seven years and to all who encouraged me to pursue this exciting path of doctoral study. First, many faculty members have been of great assistance on my journey to and through my doctoral degree: they coached, formed me and challenged me constantly. I have learned much about the professionalism of philosophy from one of the best philosophers and I have significantly improved my philosophical perspective while being at UCSD. Second, the brilliant graduate students and the engaging intellectual life of our community made the dissertation process much easier and more pleasant. Many colleagues, friends and audience subjected themselves to talk about my projects and my ideas even at incipient stages. I want to mention here among others: Anna Alexandrova, Matthew Brown, Andrew Hamilton, P.D. Magnus, Tarun Menon, Robert Northcott, Aaron Schiller, Jacob Stegenga who provided me with deep insight and constructive criticism. Now I feel deeply influenced by the UCSD style of doing philosophy. My UCSD years were by far the best of my life and I am sure that my future scholarship will reflect the mentoring I benefited from all the colleagues and friends at UCSD.

During all these years I have had the fortune to be part of a thriving community of philosophers of science and philosophers of physics. I was fortunate enough to see how the community of philosophers of physics in Southern California has grown steadily over the last seven years. I benefited from the invaluable help of the friends and colleagues who participate in The Southern California Philosophy of Physics Reading Group. I am proud to be part of this thriving community of philosophers of physics. I want to thank especially Jeffrey Barrett, David Malament and Christopher Smeenk who provided valuable comments and criticisms to different parts of this project. I am proud to see that UCSD, UCI and UCLA are gaining strength in philosophy of science and philosophy of physics.

Last and surely not least I want to thank my wife Susan for her steady intellectual support. I feel terribly lucky that I met her. Without Susan this project would look poorer and less animated. Susan has always been supportive to my academic career, including polishing and improving the present work. She always pressed me for the practical side of my research and for immediate consequences of it. If I have not found one, it's my entire fault. Her question: "Can you change something in the world with your dissertation?" still haunts me. She corrected and asked for clarifications for almost all pages of this thesis. Her brand of academic work is done within a skeptical stance toward anything that is too "good to be true". She also introduced me to political science and to the wonderful world of social science methodology. I wish I would know more about the philosophy of political science!

For over four years Susan found a way to be my best partner and best friend, even when the pace of my academic progress was sluggish. Without her support this project

would not have been accomplished. I would like to thank my family in Romania for their support and for believing that all I did was right even when it was not. I cannot express how much it has meant to me over the years. I thank my parents for supporting my vision and helping me fulfill it.

Before coming to UCSD I was interested in the methodology of science and what makes science a special type of knowledge. It would not be fair to fail to mention the names of the teachers I have at the University of Bucharest: Professor Ilie Pârvu and Professor Mircea Flonta had a major influence on me in the early years of my education in Romania in a period of turmoil and confusion. They both helped me in shaping my philosophical career in the direction of philosophy of science and encouraged me to explore philosophy by using my background in physics. I had great colleagues at University of Bucharest and many of them inspired me in my journey. I benefited of three months of intense and fruitful instruction at CEU (Budapest) under the supervision of William Newton-Smith where I had been introduced to some topics in the philosophy of science such as:laws of nature, rationality of science, dispositions and non-empirical virtues of scientific theories. This dissertation originated in 2003 after discussing with Craig the dimensionality of physical bodies and of spacetime. Later on, I participated in a reading group on scientific unification in which Mark Newman, Andrew Hamilton and Eric Martin were the main actors. I thank them all for being part of this project and inspiring me with their skepticism. Starting from Morrison's book I decided to follow uncharted paths in philosophy of science by analyzing an episode in the early evolution of String Theory after taking an exciting class on String Theory in the Department of Physics taught by Ken Intriligator. Other friends and colleagues have helped me with feedback and suggestions and I mention only some of them

here: Sorin Bangu, Matthew Brown, Valentin Cioveie, Sorin Costreie, Magda Dumitru, Koray Karaca, Douglas Kutach, Laurian Kertesz, Dennis Lehmkuhl, Tarun Menon, Ilie Pârvu, Andrew Wayne, and Steven Weinstein.

The editor Dennis Dieks and the anonymous referee of the volume *The Ontology of Spacetime* published by Elsevier provided valuable remarks and corrections on the accompanying paper (Muntean 2008, 279-299). The same goes for the anonymous referees of the *Revista Română de Filozofie Analitică* where I published a related material.

The *Second International Conference on the Ontology of Spacetime* (Concordia University, 2006), several meetings of the *Southern California Philosophy of Physics Reading Group* (University of California, Irvine, 2005-2009), the *Philosophy of Science Retreat* (at the James Reserve in the San Jacinto Reserve, 2008 and 2009) and the *Geneva Summer School in the Philosophy of Physics* (Arolla, Switzerland, 2008 and 2009) were all wonderful opportunities to discuss spacetime theories. I want to thank the organizers and the participants for their patience in discussing exotic spacetime theories.

All these nice memories are present in the pages of this dissertation.

VITA

Specialization: Philosophy of Science, Philosophy of Physics, Metaphysics
Competence: Cognitive Sciences, History of Science, Philosophy of Technology and of Artificial Intelligence, Modern Philosophy, Logic, Ancient Philosophy

Education

University of California, San Diego — Department of Philosophy

2009 Ph.D. in Philosophy

University of Bucharest — Department of Philosophy

2005 Ph.D. in Philosophy. Specialization: Metaphysics. Chair: Ilie Pârvu

1997 M.A. in Theoretical Philosophy. Grade: 10 (Eu 10-point scale). Chair: Ilie Pârvu

1996 B.A. in Philosophy. Grade: 9.88 (Eu 10-point scale)

Polytechnic Institute of Bucharest - Department of Communication

1992 B.S. in Electrical Engineering. Specialization: Applied Physics. Grade: 9.12 (Eu 10-point scale)

Selected publications

2008 Mechanisms of Unification in Kaluza-Klein theory. In Dennis Dieks (ed.), *The Ontology of Spacetime, II*. Elsevier, Amsterdam, 275-300.

2007 Autonomous Agency, AI, and Allostasis: A Biomimetic Perspective. Co-authored with Cory Wright. *Pragmatics and Cognition. Special Issue: Mechanicism and autonomy: what can robotics teach us about human cognition and action?* 15 485-513.

2007 Leibniz and the Mathematization of Forces. *Revue Roumaine de Philosophie* 51 100-120.

2005 Philosophy of Science: Laws. Co-authored with Anna Alexandrova; Nancy Cartwright; Sophia Efstathiou; Andrew Hamilton. In Frank Jackson and Michael Smith (ed.), *The Oxford Handbook of Contemporary Analytic Philosophy*. Oxford University Press, Oxford, 792-818.

Academic Employment

University of California, San Diego

2003 – 2008 Graduate Student Instructor — Department of Philosophy

Spring 2009 Associate Lecturer (temporary appointment) — Department of Philosophy

University of San Diego

2006 Adjunct Professor (temporary appointment) — Department of Philosophy

University of Bucharest, Romania

1999 – 2007 Associate Professor (tenured 2000) — Department of Philosophy

1997 – 1999 Assistant Professor (tenured 1998) — Department of Philosophy

Polytechnic University of Bucharest

1992 – 1996 Assistant Professor (tenured 1996) — Department of Social Sciences

A full version of my Vita is available at: http://imuntean.net

ABSTRACT OF THE DISSERTATION

Unification and Explanation in Early Kaluza-Klein Theories

by

Ioan Lucian Muntean

Doctor of Philosophy in Philosophy

University of California, San Diego, 2009

Professor Craig Callender, Chair

Unifying distinct domains of phenomena is one of the most important non-empirical virtues of scientific theories. However, what *counts* as unification and what makes it *important* are philosophically controversial. I canvass two positions toward unification (the enthusiasts and the dissenters) as well as two methods to approach unification: the general approach and the specific approach based on case studies. Some philosophers take unification to be truth conducive (Friedman, Glymour, etc.) others to be

central to scientific explanation (Kitcher) and still others find it to be typically neither (esp. Morrison). To make progress on these questions, attention should be paid to concrete, historical episodes.

In my dissertation I tackle one of the most significant episodes in the history of physics, an episode that—oddly given how important the theory is now in the context of String Theory—has escaped historical and philosophical investigations or it has been under-investigated. That episode is the early attempt to unify gravity and electromagnetism within a five-dimensional spacetime by Kaluza (1921) and Klein (1926). This theory is philosophically interesting in its own light, but as the ancestor to current attempts to unify gravity with matter fields, it is rich with consequences for the contemporary foundations of physics.

Morrison (2000) argues that many instances of unification are trivial, spurious or related neither to explanatory power, nor to scientific realism. Others have recently argued that unification in general is neither necessary, nor sufficient for explanation, although there may be some (weak) correlations between unification and explanation. Against this background I emphasize the novelty of my approach by making room for a new type of unification illustrated by Kaluza, and especially by Klein, in which unification is strongly related to explanation. Although some aspects of my case study are suggested in the philosophical literature, they are never fully discussed.

I argue that, as a two-stage process from Kaluza to Klein, the Kaluza-Klein theory brings about an increased unificatory and explanatory power and becomes less *ad-hoc*. Kaluza's theory is interesting because it is, arguably, almost a real-life case of a spurious unification (save his speculations about quantum mechanics). Klein improves significantly on Kaluza and proposed a curled fifth axis (a procedure called “compactification”), explains the quantization of electrical charge, uses fewer brute facts and fewer types of symmetry, and solves problems Kaluza could not.

As the five-dimensional theory became more unified with Klein, I argue that it has a greater explanatory power. In addition, I show how the sense in which Klein’s theory is unificatory is interestingly different than in some other unificatory theories (in contrast to e.g. electromagnetism). Unlike Kaluza, Klein employed an extrinsic factor: the behavior on the fifth dimension of a wave-function—present neither in gravity nor in electromagnetism—which has had its own interesting history.

Kaluza-Klein offers a novel type of unification; Klein’s unification, in particular, constitutes a type of unification which is neither reductive, nor synthetic. In opposition to some dissenters, I show in greater detail how unification works in the practice of science and how it relates to explanation, simplicity, theory validation, etc. I claim that the recurrent skeptical positions are rooted in a misunderstanding of both the concept of unification and the concept of scientific explanation.

Finally, I stress the importance of the Kaluza-Klein type of unification for recent attempts to explore extra-dimensions of spacetime (related mainly to String Theory).

PART I. PHILOSOPHICAL PERSPECTIVES ON UNIFICATION

Chapter 1. Arguments for unification

Although unification has constituted a hot topic in the last three decades, both in the scientific and philosophical literature, philosophers as well as scientists have mixed attitudes towards it.[1] There are enthusiasts for unification and lots of skeptics. Skeptics claim that unification is hard to define, difficult to achieve and often not worth the price. And even if one manages to define or delineate it, unification is hard to achieve. Although I agree that unification is more of a goal than an achievement because in many theories it is simply not present, I side with the enthusiasts. In my case study I see evidence for unification and furthermore I argue for its role in the progress of science. I argue for and show a strong connection between unification and explanation. I provide arguments for unification and address some of the aforementioned skeptical positions. Moreover, my dissertation pays special attention to the price paid for unification in Kaluza's and Klein's cases. Also, I adopt a comparative method: by weighing the drawbacks against the advantages gained, I argue that Klein's theory is more unificatory, more explanatory, and finally a better theory than Kaluza's in several respects other than unification and explanation: it solves several problems, is less ad-hoc, has a better vacuum stability, etc.

Similar to other concepts in the philosophy of science, such as explanation or scientific realism, the status of unification and its relevance to scientific progress are fraught with controversies. What is not controversial is that unification is present in science, at least in exact sciences, and that some theories are more unificatory than other

[1] In order to avoid repetitions, in this dissertation I will intentionally omit the attribute "scientific" as I deal almost exclusively with *scientific* unification. Other meanings of unification are used accidentally and are identified explicitly.

theories. A quick survey reveals that unification is a major recurring theme in the 20th century physics. In seeking new theories not yet empirically confirmed, physicists often espouse a desire for theoretical virtues akin to unification and strive to reach it for reasons ranging from aesthetic considerations like elegance, simplicity and harmony, to more pragmatic reasons such as the scantiness of language used or of availability computational tools. Realizing that a phenomenon is not what it seems and belongs to a different class is part of the unificatory story: history of science abounds in discoveries that an odd phenomenon was a case of something more general.

Some would say that unification is easy to recognize but difficult to define: "you know it when you see it." We have, seemingly, an intuition of an economy of knowledge when much is realized with sparse resources, similar to the way living beings around us are able to optimize their existence. Some uncontested successful stories, both from the current practice of science and from its distant history in which unity of knowledge has improved bolster our intuition that unification can play a major role in the progress and practice of science. Is there a way to make this intuition of unification more precise?

At a first sight, unification is akin to scientific reduction, although they are not the same thing. In a famous paper, P. Oppenheim and H. Putnam claimed that the unity of science in the strongest sense was realized if the laws of science were not only *reduced* to the laws of one discipline, but the laws of that discipline were *connected* and *unified* (Oppenheim and Putnam 1958 4). The two authors then confessed that they never figured out how this requirement could be made precise.

Take a simple case of reduction. The Stefan-Boltzmann (circa 1880) theory of radiation that showed the dependence of the energy density with the fourth power of temperature needed a constant of proportionality, κ. By using the quantum hypothesis, Planck showed in 1900 that κ can be defined in terms of the speed of light, Planck's constant, Boltzmann's constant, etc. Later on, other theories provided connections between these free parameters of Planck's theory. One can see how and why the theory of radiation was finally reduced to quantum mechanics. Fewer constants and fewer theoretical terms explained more phenomena than they did before. Does this case constitute unification or is it a mere reduction? I deal with similar questions in my case study, where I argue that we have unification where reduction is not possible or not the optimal solution.

In many other cases the relations among theories are far less obvious. P. Teller expresses the uncertainty surrounding unification in a concise way: "I agree that unifications [and reductions] show something important about how our theories bear on the world. But I take the worries to show that we are very far from understanding what that 'something' is" (Teller 2004 443). P. Lipton also saw the main difficulty of the unification account of explanation—roughly, we understand and explain a phenomenon only when we see how it fits together with other phenomena—in the fact that "the notion of unification turns out to be surprisingly difficult to analyze" (Lipton 2004 28).

1.1. Some possible definitions of unification

What about the definition of unification? In the last decades there were some attempts to define unification as formally as possible. I will discuss two of them here.

Watkins (1984). John Watkins tried to define unification *à la* Lakatos. For Watkins, a theory T is composed of a theoretical core H (called also "metaphysical core") of axioms that contain only theoretical predicates and a set of auxiliary assumptions A that are a mixture of theoretical and observational predicates.

Watkins proposes to use a Ramsey-sentence T_R to separate the theoretical part from the empirical part of a theory. The theoretical core H is then the set of sentences which are not consequences of T_R and they are not testable (Watkins 1984 194). A theory $T_j = A_j+H_j$ is more deeply unified than a competitor T_i when its theoretical core more effectively increases the testable content: the testable content of T_j minus the testable content of its auxiliary assumptions A_j are greater than the testable content of T_i minus the testable content of A_i.

Watkins started from the level of systems of theories that unifies their components: Let S_i and S_j be theoretical systems consisting of the conjunction of a number of theories, where S_j has superseded S_i. Assume that S_j has at least as much testable content as S_i. Watkins here uses the Popperian concept of comparative testability, CT, and writes that $CT(S_j)>CT(S_i)$. Then we can say that the progress from S_i to S_j involves greater unification if the number of unified theories in S_j is less than in S_i. An example is when S_j is the Newtonian theory and S_i is Galileo's and Kepler's laws (Watkins 1984 215).

In my opinion, Watkins's account is fundamentally flawed: first, there are very few scientific theories which are axiomatized (even if some field theories are axiomatized, the meaning of axiomatization does not fit Watkins' description). Second, even if some areas of theoretical physics can be axiomatized, there are several, non-trivial ways to achieve an axiomatization in which the domains H and A are not the same and many of them are not even remotely similar to the way arithmetic was axiomatized by Peano (Wayne 1996 393). Despite Watkins' efforts to provide rules of the naturalness of axioms, a difficulty in deciding which axiomatization is the most natural still lingers; see (Watkins 1984 208) but also (Glymour 1980 39). The practice of theoretical science proofs that the naturalness of an axiomatic system is not a well posed philosophical problem: axiomatization is always a tricky business for the simple reason that there are too many ways of axiomatizing a body of knowledge. Moreover, the domains *H* and *A* change constantly in the history of a scientific theory and any attempt to decide one for good the distinction H versus A is hopelessly difficult. Last but not least, testability is theory-laden. Many attempts to characterize theories by translating them into a simplified, formalized language, are doomed to fail in almost all cases. With some notable exceptions, people gave up axiomatizing scientific theories. I do not tackle axiomatized general relativity or electromagnetism, although there were successful attempts to axiomatizing them both.

What is missing in all formal approaches to non-empirical virtues of theories (along with unification, simplicity for example was formally discussed by N. Goodman, M. Friedman, P. Kitcher, and J. Watkins *i.a.*) is the content of the theory and not its structural or formal components. I adopt here an account of unification sensitive to content. I do not want to speak about a material unification, but the suggestion is that we need far more content than Watkins suggested. Exaggerating the formal as opposed to the material is neglecting the fact that the syntax is used in science to relate content. In the 1930s several logicians (G. Gentzen, S. Jaśkowski) suggested that even in logic where formal procedures reigned, the meaning of the logical operators could be gathered from the rules that governed their use in inference. The dynamics of meaning and significance of theoretical terms casts serious doubts upon axiomatization and consequently upon a formal approach such as Watkins's. Unlike mathematical theories or logical systems, scientific theories are constantly changing. For the present analysis of unification, axiomatization plays almost no role, *pace* Watkins. Unification worked pretty well even when axiomatization was not present or it would have been discovered much later.

Strevens (2004). A different approach is to look for some concepts of unification that capture some or the most features of the relevant instances of unification available in the history of science. If unification in itself is difficult to define, some think that it can be related to other non-empirical features of theories such as: simplicity, parsimony, generalization, identification, integration, etc. In fact, there are some answers available. An effective way is to define a unificatory theory as instantiating a good balance among "some" theoretical features. For example, M. Strevens bases his definition on three concepts frequently mentioned in the literature—generality, simplicity and cohesion:

> The unifying power of a theory increases in proportion to the following properties of the theory:
>
> - *Generality*. The number of actual phenomena that can be derived using the theory,
> - *Simplicity* […][2]
> - *Cohesion*. This third desideratum has been characterized in a number of ways. The aim of the desideratum is to discriminate against theories that, rather than picking out real patterns of phenomena, pick out mere unpatterned conjunctions of phenomena (or perhaps even worse, all possible phenomena) (Strevens 2004 155).[3]

I see several problems here. Strevens moves the difficulties on block down the road: simplicity and cohesion are in themselves difficult to define. Generality of two theories is difficult to compare when the phenomena described are totally different and

[2] Another author who takes simplicity as part of the unity of a theory is N. Maxwell. As I admitted from the beginning, I prefer to draw a sharp distinction between unity and unification. See (Maxwell 1998 157).

[3] A cohesion principle is formulated by Kitcher as the "stringency" requirement. For others, cohesion is related to the number of parameters or set of parameters used by a theory in explanations.

hinges upon the procedure of counting phenomena. All theoretical virtues have multiple meanings and were both applied to the knowledge as a whole, to the world, to the language, to the totality of our experience, etc. As in the case of empirical virtues, almost all theoretical virtues of a theory are controversial: they can be trivialized or, on the contrary, can become too high a standard for theories to achieve. Many physicists look for a theory of everything (**TOE**) that is as simple as possible: Leon Lederman, the director of the Fermilab claimed in the late 1980s that the ultimate formula of the **TOE** is that simple that "you could wear [it] on your T-shirt" (Davies and Brown 1988 7). Although this is a very colorful description of a **TOE**, questions related to simplicity remain: How do we count facts that can be derived from a theory? How do we define simplicity independently of the language in which the theory is formulated? What is a pattern of phenomena? What is the relation in a pattern? Several authors proved that simplicity is relative to language: any complex theory can be transformed into an ostensibly simple theory by an appropriate change of language. As is clear from an elementary study of the history of scientific progress fixing a language is not possible. In many cases the language is completely discarded during a scientific revolution and theories become incommensurable (Maxwell 1998 38, 157; Bunge 1963); in those case simplicity and generality need to be completely redefined. In some respect, the definition begs the question because simplicity and generality are defined in terms dependent on what unification is. We will see that Morrison and Woodward expressed similar worries. In this dissertation, I do not *define* unification by stipulating necessary and sufficient conditions, but I attempt to *delineate* it by analyzing its instances. I also start from some philosophical positions about unity and unification as well as from what unification means in science.

Are there sufficient conditions for unification that can work in all cases? Most likely there are no such general sufficient conditions and we may be contented by delineating unification instead of defining it.

1.2. Unity of a theory and unificatory theory

In order to clarify the concept of unification, I introduce a preliminary distinction between the *unity of a theory* and the *unificatory theory*.

(A) Unity of a theory. One can argue that *unity* is a feature of *one* theory, call it "theoretical unity". It is frequently said that a theory T is *unified* when T predicts and explains a large class of phenomena by using a relatively small theoretical structure. This is perhaps a very general concept and unfortunately too ambiguous to be used in a philosophical context. What is a small theoretical structure? What do we compare it to? Just to anticipate the discussion to follow: M. Friedman's suggestion was to count on the brute facts which the theory relies; J. Watkins elaborated this view and proposed a schema based on law-counting; in criticizing Friedman, Ph. Kitcher proposed to count the set of patterns of arguments used by the theory. If one embraces the idea that a scientific theory possesses an ontology, one can think that a theory is unified when, given a particular interpretation, it is *ontologically parsimonious*, i.e. it has a small "ontology" of few elementary entities that explain lots of phenomena (Baker 2003; Nolan 1997; Ducheyne 2006). This was arguably the ideal of the Standard Model (Wayne 1996 395). For example, what matters for Wayne is the *prior ontological commitments* of the standard model and not the formal derivations. "Gauge invariance presupposes commitments to physical forces producing the dynamics of the subatomic domain" (Wayne 1996 403). But counting facts, laws, patterns of argu-

ments, or the elementary entities used by a theory are all vexed questions (Watkins 1984; Wayne 1996; Friedman 1974; Kitcher 1981). It is enough to mention here that "counting" is relative to an interpretation, to a language or to a specific formulation of the theory, to the class of the models the theory can have, etc. One of the philosophical interests in non-empirical virtues of theories such as: unification, explanation, elegance etc. is to analyze their dependence on the language in which theories are formulated. We can concoct very elegant theories that look horrible in a different language: the same can be said about unification. As I show, trivial and spurious counterexamples of unification are handy for all these definitions.

(B) Unificatory (or unifying) theory. In my dissertation I deal with a specific case of unified theories that I call hereto "unificatory" (or "unifying") theory.[4] It is not infrequent that a scientific theory T unifies a range of phenomena that were previously described by two distinct theories, T_1 and T_2. First of all, T is considered a "new" theory, albeit it does not make *new* predictions as compared to T_1 and T_2. What does a scientific theory do besides providing predictions? There are several answers available: a scientific theory (a) explains facts about the world and/or (b) represents/describes facts.[5] For many (a) suffices (L. Sklar); for others, (b) is really the central aim of scientific theorizing (W. Salmon, J. Woodward, P. Kitcher). For the majority of philosophers, both (a) and (b) characterize science and possibly other aim besides (a) and (b): let us call it X. Intuitively, T is more general than T_1 or T_2 taken together because it can represent/explain/X more facts than T_1

[4] For the present purposes, the two terms are interchangeable.

[5] For the moment I do not need to fancy a theory of scientific representation. One can embrace the semantic view of the theories: Giere's relation of fitness or similarity of the model M of a theory and the world would be enough here (Giere 1988). Another candidate for the relation of "representing" is partial isomorphism (French 2003, 1472-1483). A critical discussion of this relation is found in (Callender and Cohen 2006).

or T_2, in a more economic way. Let us clump together all these relations in a general predicate **A** that describes a composite relation of applicability of a theory T to a fact or set of facts *d*: explaining, representing, describing, increase the knowledge about, etc., can be all parts of the *applicability*. Another thing: the number of facts necessary to represent a specific domain can be called the "free parameter" of the theory.

Many authors conflate (A) and (B). But it is an empirical question whether a unified theory is unificatory or vice-versa. I plan to focus here more on (B) and to discuss the unification and the details of unificatory theories. I do not suggest that there is a dichotomy between (A) and (B) above; although my preference goes with the latter: I take unification as a process of theory-creation and of theory-choice, not as a feature of one given theory. Moreover, by scientific unification I do not mean a general relation between parts of science. I prefer to discuss scientific unification as a connection between *theories*. A different concept of unification, more general than the one used here, is the unification of scientific fields instead of theories by *interfield* relations (Darden and Maull 1977, 43-64). Another, even more general, concept of unification would be one that discusses the unifications of disciplines within one science or even the unification of two sciences. What is important is that all these approaches to unification are anti-reductionistic in nature. In general the dis-unity movement does not dismiss local attempts at unification, but finds unification at the "global" level unconvincing (Dupré 1993 228)

.

My approach remains local in all these respects. Because my case study has its own specificity, I deal especially with (B) in my dissertation. I do not dismiss the importance of (A), but in general, (B) covers more tangible cases of unification in science and unification refers preponderantly to (B). Moreover, saying that a theory uses a small number of X (e.g. laws, facts, patterns, member of an ontology, etc.) is already a relative concept. I have a strong preference to treat unification in the context of scientific holism and to relate T to previous, existing theories. In very few cases, previous “theories” were mere collections of pre-theoretical intuitions or outlooks: in many situations, they were full-fledged, mature and respectable scientific theories. I do not endorse the idea that unification occurs only after the unified theories acquired a certain level of “unity”—that would make (A) a condition to achieve (B). Let us take (A) and (B) as two independent definitions.

In order to simplify the terminology, I refer to unification as a chain of procedures, which has as a result, the discovery or creation of a unificatory theory. Unification is more or less a process of creating new scientific theories and in my opinion it is directly related to scientific progress and scientific discovery.

For my current purposes, here are some conditions that characterize unification:

Def 1 Given two theories T_1 and T_2, a third, different theory T, is unificatory:

(I)If data d_1 is in relation **A** with T_1, d_2 is in relation **A** with T_2, then d_1 and d_2 are in relation **A** with T.

(II) If T_1 has n_1 free parameters and T_2 has n_2 free parameters, then T has $n < n_1+n_2$ free parameters.

(III) If T_1 explains m_1 facts and T_2 explains m_2 facts, then T has explains $m > m_1+m_2$ facts.

etc.

I am not particularly interested in (I) here because in the case of both Kaluza and Klein there are few data to be accounted for by *T* besides was T_1 and T_2 already display; in fact, my whole dissertation will be centered on (II) and (III).

First, a very general remark: for space reasons and for the specific episode on Kaluza-Klein theories discussed below, I prefer to adopt the syntactic characterization of theories. But this choice is not central and can be easily translated into a different language. Similar formulations could be generated for the semantic views of theories. For example, if one prefers to use the semantic approach rather than the syntactical view of scientific theories, then one can express T_1 and T_2 as "classes of models", Δ_1 and Δ_2 instead of collections of sentences. One needs to replace (I) in **Def 1** with: for every model $M_1 \in \Delta_1$ and $M_2 \in \Delta_2$, there is a model *M* such that both M_1 and M_2 are embedded in *M*; for a detailed view on the structuralist unification, see (Bartelborth 2002 98-103; Friedman 1983).

Definitively, (III) constitutes the main thrust of the discussion on unification. For several reasons to be discussed further, I include explanation in the set of relations **A**. Indeed, here explanation has a special role to play. Describing or representing the world simply does not suffice. *T* enlarges the explanatory store of the conjunction of T_1 and T_2 by using fewer free parameters. Morrison has argued that the unifying formal structures—let us say, the gauge invariance or renormalizability—present within the Standard Model "embody the greatest possible generality" as all patterns of arguments used by the Standard Model share a small amount of formal structure (Morrison 1995 16-17). It also can be less idealized, simpler, more beautiful or elegant or stronger compared to the previous theories, but this is not by any means necessary. If generality, simplicity and explanatory power are virtues of theories, so is unification. If a new theory is simpler and more beautiful than the previous theories, so much the better for it. But what is central here is the promise of T to *explain* more with fewer parameters compared to the previous theories T_1 and T_2. I suggest that among the benefits a unificatory theory brings about, explanation has a central place. In the last years explanation is heavily discussed in philosophy especially in the context of causation (Strevens 2008; Campbell, O'Rourke, and Silverstein 2007; Salmon 2006; Salmon 1998; Psillos 2002, 324). My dissertation is about explanation is the context of unificatory theories.

There are some successful episodes in the history of science that spawned unificatory theories. Newton unified terrestrial and celestial phenomena under one theory; Maxwell unified electricity and magnetism; Joule unified heat and mechanical energy; Darwin's theory of "descent with modification" described by only one principle the biogeography, comparative anatomy, embryology of virtually all living beings; Einstein unified electromagnetism mechanics with the Galilean relativity of mechanical motion; subsequently space and time, etc.[6] Mendeleev discovered in the 1870s the periodic table of the seventy or so elements: he showed that that was a link among all possible atoms. His theory was much more unified than the previous chemical theories that simply classified them in an unsystematic way. Less intuitive and less elegant, but more spectacular, is the Standard Model of elementary particles. It is a unificatory theory of three interactions under one and the same formalism.

[6] On the other hand, natural selection played the explanatory role in Darwin's theory. Thanks to William Bechtel from clarifying this point.

Perhaps both (II) and (III) in **Def 1** are problematic because in general it might be difficult to count facts and parameters.

1.3. Two attitudes towards unification

There are controversial aspects of all these exemplar unifications. First let us ask a simple question: is unification always grounded on two existing theories? My answer is: not *always*. Some unificatory theories do not unify two previous theories in a clear way in the sense of my **Def 1**. Cases of unificatory theories in sciences other than physics, although more controversial, are worth mentioning. For example, although Darwin used Malthus' theory and Linnaeus' taxonomy in shaping his evolutionary theory, it is too much to say that evolutionism unified Malthus and Linnaeus. In Darwin's own words, because the evolutionary hypothesis explains "various large and independent classes of facts" it is unificatory and moreover "rises to the rank of a well-grounded theory" (Darwin 1868 12). Perhaps under Whewell's influence, Darwin saw unification in explanatory terms; he also inferred from the best explanation to the truth of his theory.[7] But it is not clear at all whether Darwin intended to unify two theories or he simply used them in the process of unification. Moreover, the synthesis of organic and inorganic chemistry achieved in the mid-19th century (F. Wöhler and R. Woodward pioneered a unification of organic and inorganic theories within chemistry);[8] the explanation of all properties of chemical elements by the quantum mechanics of the electrons orbiting the nucleus composed of protons

---

[7] "It is the consideration and explanation of such facts as these which has convinced me that the theory of descent with modification by means of natural selection is in the main true." (Darwin 1868 14)

[8] A critical discussion of a later episode in organic chemistry, the so-called "Ingold revolution", and some interesting conclusions about explanation and unification in chemistry can be found in (Goodwin 2007)

and neutrons;[9]theoretical population geneticists working within evolutionary synthesis provided a unificatory theory of how Darwinian evolution is possible within a Mendelian system of inheritance (A separate unification, although not by the way of a theory, is Linnaeus' biological classification. (Plutynski 2005; Morrison 2006 ch. 6); various theories and methods were proposed in order to unify psychology and confer it a scientific status (the "unified positivism" of A. Staats; for a philosophical critique of these proposals, see (Kukla 1992); Chomsky's hypothesis of an innate language had been interpreted as unificatory (See also Chomsky's own thoughts about unification in (Chomsky 2000 esp. ch 5)). Needless to say, all of these cases are more or less controversial and they are only loosely harnessed by the aforementioned intuition of unification.

A skeptic may ask: so what? Is there something special about unification? Do we aim for unification as we seek other virtues of theories such as empirical success, confirmation or prediction? Is unification accidental, or is it intimately related to how sciences work? At one extreme, some philosophers endorse wholeheartedly unification as strongly related to the way science progresses. Kitcher's unificationism, for example, is usually read in the spirit of saying that we should look for unifications and that unification is a major achievement when we find it. At the other edge, others would demote unification as mere mathematical trickery or as a mere psychological illusion stemming from our way of representing the world. In many cases, the dissenters blame mathematics for creating the illusion of unification and then drawing us into this illusion further by entertaining it when

[9] In this respect the virtual unlimited types of substances were reduced to four elements: electrons, protons, neutrons and photons. It is still a good question whether chemistry can be reduced to physics or whether quantum chemistry is a unification of chemistry and physics. For recent debates around this issue, see (Hendry and Needham 2007). For example, are the properties of the orbital absolutely central to quantum chemistry, or are they *totally* reducible to quantum mechanics?

it is not genuine. We have the illusion that under the formal part of mathematics, which can treat almost everything in the same way, there is unification. Mathematics is unreasonably effective in creating unificatory theories. Some skeptics say that there is no unification where there is no mathematics (Morrisson *i.a.*): that means there is no unification in the absence of a mathematical formalism? What about unification in biology?[10] The same skeptics ask for a further connection between the mathematical machinery and the real world. In commenting on the case of electroweak unification, Morrison echoes this attitude:

> The process of unifying these forces was driven by considerations grounded in the mathematics of gauge theory, rather than in the phenomenology of the physics. Hence, in this case, to an even greater extent than in the others, I consider unity to be a product of the mathematics, rather than a verification of a detailed causal hypothesis about relations between diverse phenomena or natural kinds. The result is a unified theoretical framework that integrates forces that, at the level of phenomena, remain ontologically distinct (Morrison 2000 109-110).

As a middle way, some see unification as an ideal, “a bridge too far” to reach that comes with a destination that carries too high of a cost, albeit not being mere trickery or a delusion. But for the skeptics, for all practical purposes, the normative aspect of unificationism is not relevant.

In the light of these controversies, it behooves one to distinguish two opposite attitudes toward unification and within them some extreme and moderate positions. On one hand there are the enthusiasts, both philosophers and scientists, who always press for unification. On the other hand, a large group of dissenters question the centrality of unification for the progress and practice of science. Among their reasons the most frequent are:

[10] Morrison suggested that unification in biology had occurred within statistics (Morrison 2000 ch. 7).

unification can be easily trivialized, the exemplar cases of unification have as counterparts some blatant counterexamples, some unifications are based on *ad-hoc* assumptions, unification in special, non-exact, sciences is difficult to find, etc.[11] For Kitcher that both groups are absolutely necessary to balance the scientific search for unification:

> If we think of unification as a regulative ideal for a scientific community, then the best way of approximating the ideal might be to have two kinds of people, those always pressing for unification and those always insisting on the particular details, each group keeping the other honest. Nancy [Cartwright] and John [Dupré] would be wonderful representatives of one of these groups (Kitcher 1999 343).

Those who press for unification are usually outnumbered by the dissenters or the detail-oriented philosophers. Kitcher tries to suggest that details drive us away from unification, i.e. from the “big picture”. The present dissertation can be read as “pressing for” an analysis of unification, but based on the “details” of individual case studies, as it were.

1.4. Strong philosophical positions about unity and unification

The partisans of unification who press for unification everywhere is science draw inspiration from great successful unifications mentioned above and have repeatedly claimed that unification is a great virtue of a theory. Authors such as W. Whewell, C. Hempel, M. Friedman, P. Kitcher, argue from various positions and using different premises *for* scientific unification. M. Friedman held unification in such high esteem that he believed unified theories were more likely to be true than the dis-unified ones. This is to say that unification is truth conducive. As such, it can play a crucial role in an argument for scientific realism. At this end of the spectrum you can find the metaphysicians who

[11] It might have been the case that unification has worked work in some areas better than in others. I will briefly discuss some instances of unification in biology as well as the possibility of having unification within mechanistic philosophy although these are not among the major themes of my dissertation. See the discussion about sectors in the metric of Kaluza and Klein, in the Section 13.2 and 14.1.

attribute a deeper significance to scientific unification that goes beyond science. Here I make a distinction between metaphysical and epistemological aspects of unification.

The authors discussed in this thesis, M. Friedman (mainly in his 1974 paper), P. Kitcher and M. Morrison, approach unification from an epistemological perspective, rather than metaphysical. Unification is taken as an epistemic virtue of a scientific theory. P. Kitcher finds scientific explanations that unify so powerful and pervasive that he is inclined to more or less identify unification with scientific explanation: to scientifically explain some phenomenon is to embed it in a unifying pattern of argument. Consequently, scientific explanation is defined in terms of unification.

I see the metaphysical approach to unification as an argument pertaining to show that as a feature of our scientific theory unification relates to something in the world. Think of an analogy: causal accounts of explanations claim that there is something out there to explain, i.e. the causal structure of the world. The metaphysical unificationist claims that there is a direct referent for unification in the real world, so there is metaphysics behind unification. Friedman (1983), Whewell and other enthusiasts of unification seemed to suggest that unification has a referent "out there", yet to be discovered. The candidates for such a reference are:

(a) a metaphysical unity of the world;

(b) the causal structure of the world;

(c) for a given theory, the existence of single mechanism that produces a variety of outputs.[12]

[12] I include here the mechanistic approach to unification.

Concerning (a), the concept of scientific unification can be linked to the philosophical idea of unity of nature which served as central regulative ideals in philosophy. Various Ancient monistic philosophies explained the whole world from a sole principle and dis-unity was apparent. A monistic doctrine wholly explains the world from one principle without questioning or being concerned about knowledge. These rich philosophical ideas are all more or less remotely related to the concept of unification I envisage here. F. Bacon, R. Descartes, G. W. F. Leibniz, I. Kant and some of the Positivists of the 19th century (most notably E. Mach) hinted toward the idea of unity of the sciences within one theoretical framework.

The rationalist argument is as follows: If science is an exemplar of knowledge, and if knowledge is ideally unitary, then science should be unitary, as well. Leibniz illustrates the doctrine of unity of science made possible by the universal science in the Rationalist tradition at its best. He envisaged this universal science as "algebra", although he admitted that nobody had ever used or suggested it. Once we have apprehended the mathematical method and we have depleted science of its content, i.e. numbers or geometrical figures, the "algebra" can be applied to any particular science. For Leibniz and other Rationalists, disunity is conventional and arbitrary, and unity is reality. The division of science in disciplines, useful for the practice of science, is arbitrary because a single truth can be reordered in different ways and may have different interpretations:

> The entire body of the sciences may be regarded as an ocean, continuous everywhere and without a break or division, though men conceive parts in it and give them names according to their convenience. (*Apokatastasis Panan*, The Universal Restitution, a part of the "The Horizon of Human Doctrine" (1690); translated in (Leibniz 1951 73).

Influenced by philosophers or not, almost all scientists have taken for granted the idea that there is a *harmony* in nature or that Nature is wont to be simple and consonant to herself (Newton, F. Bacon). An interesting discussion is whether such an outlook influenced Einstein and the Unified Field Theory in the early days of relativity. According to E. Zahar, Einstein's and Poincaré's *methodologies* at least were influenced by Meyerson's Unity Principle: "all phenomena should fall under one all-embracing law", be it a unique geometry or a single, unique principle, such as the Relativity Principle. From here, Meyerson would infer that nature does not split into disjoint domains subject to different laws (Zahar 1980 10sqq.; Zahar 2007 152). Other philosophers and scientists looked for a unity of science at the level of language and standards: the vast majority of logical positivists within the Vienna Circle postulated the unity of science as a unity of language and method, more or less inspired by Diderot's *Encyclopédie*.

Early Modern philosophers used this ideal of unity in a more moderate way as a unity of method of sciences. Francis Bacon postulated a unity of method in science as a form of organization of empirical data. Descartes, akin to Bacon, discussed an ideal of the scientific progress as cast in terms of one unified science that at the end of the day has theories that are increasingly unified, totally predictive, absolutely exact and explanatory. Later on, some philosophers of the Enlightenment admitted that the unity of science was not necessarily related to the "unity of nature".[13] Similarly, for Kant, the unity of science is

[13] For example, in his debates with Maupertuis and Gassendi, Diderot frequently argued that we may not know what the unity of nature was, if there was such a thing, although we could and we needed to achieve a unity of science. Diderot warned in the article on the *encyclopedia* of his *Encyclopédie* (1755): "The universe offers us only individual beings, infinite in number, and virtually lacking any fixed and definitive division […] What then if the machine is in every sense infinite; if we are speaking of the real universe and the intelligible universe, or a work which is like the imprint of both? Either the real or the intelligible universe has

not directly related to a "unity of nature", if such a thing exist, but it is founded in the unity of reason. He suggested in the Appendix to the "Transcendental Dialectic" that the principle on which the unity of reason operates is as single kind of causation.[14] For a long period of time, Kant thought that mathematical exactness was the source of the unity of sciences. Science proper is mathematization: "In every special doctrine of nature only so much science proper can be found as there is mathematics in it" (*The Metaphysical Foundations of Natural Sciences*, 4:471). It is worth noting here that the Kantian idea of a unity of reason in describing the world directly influenced Th. Kaluza (see Section 10.1 for details).

A somehow extreme argument can be found in E. Meyerson's definition of scientific progress as a march toward unity: scientific explanations are all based on a natural tendency of the mind to deny diversity, plurality and change. The mind asserts the existence of constants and laws behind the transient appearances and it insists to explain the Many in terms of the One, the difference in term of sameness, and to subsume the becoming under the immutability of being. In short, scientists are all Parmenideans trying to describe a Heraclitean world (Meyerson 1962).

I see here several possibilities in the logical map of the debate around unity and disunity:

- **Strong metaphysical unity**: The unity is real and the dis-unity is apparent. Science tracks unity, but common sense is fooled by the dis-unity. The dis-unity is created by the mind or by the senses. Unification of two scien-

---

infinite points of view from which it can be represented, and the possible systems of human knowledge are as numerous as those points of view."

[14] Relevant passages are: A645-648 and B673-677 in the *Critique of Pure Reason*.

tific theories follows as a consequence from the unity of the world, although it may not be achieved immediately) and it is a direct result of the unity in the world

- **Strong epistemic unity**: The unity is created by the mind. The representation of the world is unitary and whether there is or there is not a unity in the real world does not matter. Science has unity. Unification links representations of the world, it is not metaphysical, but epistemological.
- **Strong metaphysical dis-unity**. The unity is apparent and is created by the mind and resides in our representations of the world. Unity is psychological appearance and is common in our pre-theoretical outlook of the world. Science should better live up with the real dis-unity in the world and relinquish unity. Unification may work locally only among representations of the world, but it is not related to something in the world.
- **Strong epistemic dis-unity**. The representations of the world are strongly fragmented and they are mind dependent. Science is dis-unified. Unity is apparent. With few exceptions, unification does not work and does not help.

As formulated here, all these positions are strong enough to raise suspicions and weaker positions are available. Option (b), in which unity is related to causation may help alleviate the difficulties of the stronger theses. Or one can be less ontologically committed and accept weak readings of what unification really refers to. Accordingly, a weak, epistemological interpretation of (a) is:

(a′) unification is associated to the unity of sciences,

whereas according to the weak interpretation of (b):

(b′): causation is created by the mind and unification is finding a representation of the common cause.

For some contemporary philosophers, unification is an ideal to be sought by science. In the stronger interpretation, it is related to the deep structure of the world, be it causation, disposition, or any metaphysical unity of the world. I think it is fair to mention that the very idea of relating unification to causation is in fact the oldest one. [15]

Where to place my analysis of Kaluza and Klein? The unity of the world might have been a major factor in the conceptual and metaphysical genesis of the Kaluza-Klein theories and might have acted as a "regulative idea" or, better, as a metaphor, but it had little place in the concrete development of the theory itself. Unification of two specific theories being a more local and specific enterprise, can live together with such a grand idea of the unity of the world or it can coexist with a more dis-unitary metaphysics as well. I suppose that the "strong epistemic" position is the closest to the conceptual and scientific context in which both Kaluza and Klein created their theories. But is it very useful here to discuss some stronger positions toward unification as illustrated by M. Friedman.

1.5. From unification to realism: M. Friedman (1983)

Among these overenthusiastic attitudes toward unification I stress here M. Friedman's position. He developed perhaps the strongest argument pertaining unification (Friedman 1983 ch. 6, sect. 3-4).[16] In short, the 1983 book illustrates the inference from

[15] W. Whewell's "Consilience of Induction" is an important approach to unification, maybe the first attempt to discuss scientific unification in the sense (A) but also (B), (see p. 9). Whewell's position illustrates (b′), too.

[16] This approach differs from Friedman's earlier approaches to unification to be discussed in Chapter 3 and I will explain there why I decided to discuss them in the reverse order: first the 1983 chapter and then the 1974

unification to confirmation and from confirmation to realism. The unified theory T is better confirmed than the two theories taken separately. For Friedman, only realists have reasons to believe literally in unification. Hence we should be realists in order to take benefits of this confirmation boost that unification provides. If a unificatory theory T covers two domains D_1 and D_2, and T_1 covers only the domain D_1, then T can receive a confirmational boost from both D_1 and D_2, whereas T_1 receives a boost only from D_1. From a probability point of view, T is more general than T_1, and it is better confirmed, but this means it is also less probable because there is a larger collection of possible recalcitrant data that can reject T compared to T_1.

There are several issues that must be clarified in discussing the 1983 account of unification. The specific context is the debate between relationalist and absolutist interpretation of spacetime theories. In order to make the distinction between realists and instrumentalists clearer, and to enlarge the breadth of his argument, Friedman uses two main ingredients:

- the theoretical structure $\mathcal{A} = \langle A, R_1, \ldots, R_n \rangle$
- sets of observational/phenomenological structure $\mathcal{B} = \langle B, R'_1, \ldots, R'_n \rangle$

where A, B are objects and R_i R'_i are relations. There are two options to relate the two structures $\mathcal{A}$ and $\mathcal{B}$. In the literal (or reductionistic) interpretation, B is a substructure of A, whereas in its non-literal (or representational) rendering, one has only a mapping from B to A, $\varphi : B \to A$. The realists always adopt the former, whereas the instrumentalists or the antirealists would adopt the latter interpretation. Let us say that we have a theory T and

---

and 1981 papers. The chapter holds stronger claims about unification and it is not based on the previous developments. See p. 72 for a discussion of the Friedman's two papers.

the claim that "T's empirical consequences are true" = T_E. If T is empirically confirmed, then the realist would agree to increase the probability that T is true p(T) and also p(T_E) , whereas the instrumentalism/antirealist will accept only the increase in p(T_E).

Friedman's preference for the *literal* version is obvious in the following example (Friedman 1983 240-241). If one takes object B for "gases" and for A "tiny molecules", then "by assuming that gases are literally composed of tiny molecules subject to the laws of Newtonian mechanics, we can explain the van der Waals law" (Friedman 1983 243). Once we go beyond one theory in isolation, the realist has a clear advantage over the antirealist. The realist can conjoin T with other theories to derive additional confirmation facts for T increasing p(T) beyond the original value:

> A theoretical structure that plays an explanatory role in many diverse areas picks up confirmation from all these areas. The hypotheses that collectively describe the molecular model of a gas of course receive confirmation via their explanation of the behavior of gases, but they also receive confirmation from all the other areas in which they are applied: from chemical phenomena, thermal and electrical phenomena, and so on (Friedman 1983 243).

In the specific case analyzed, the spacetime theories, the literal interpretation makes the difference between relationalist and absolutist of spacetime. It seems that in the case of spacetime theories and in the case of reduction of law of gases to statistical mechanics, representation by a map does not work and reduction fares much better:

> If $\mathcal{B}$ is literally a submodel of $\mathcal{A}$, then $\mathcal{A}$ induces theoretical properties and relations on objects in $\mathcal{B}$, properties that are in general necessary for stating accurate laws about these objects. On the other hand, the assertion that $\mathcal{B}$ is only embeddable in $\mathcal{A}$ will not induce the necessary theoretical properties and relations (Friedman 1983 240).

The theoretical structure in the case of representations is more or less a mathematical object and mapping the object from the physical to the mathematical is useful but it says nothing about the world and we have no reasons to believe in such theoretical structures. If we are free to employ any theoretical structure to derive consequences about the observables, what are finally the virtues of $\mathcal{A}$? Friedman employs two examples here: Newton's second law and Boyle-Charles law. In both case $\mathcal{A}$ contains unobservables. The structure $\mathcal{A}$ in the case of the second law is the sum of accelerations due to various non-referential systems: the rotational acceleration (Coriolis) one translational acceleration etc. all being non-observational terms (Friedman 1983 eq. 8 on p. 226). For Friedman, a pure representational structure is *adhoc*. He relates the advantages of the theoretical structure $\mathcal{A}$ to unification. $\mathcal{A}$ acquires more confirmation compared to the phenomenological structure $\mathcal{B}$ by picking up confirmation from all different areas. The pure phenomenological description receives confirmation from one area only, while the literal description gives a better confirmation: "a total theory rich in high-level structure is likely to be better confirmed than a total theory staying on the phenomenological level, even if though the latter theory may have precisely the same observational consequences as the former" (Friedman 1983 244).[17]

Friedman's idea is similar to Putnam's conjunction argument for realism (Putnam 1979). Let us say that the conjunction of two theories T_1&T_2, has as a consequence a fact E, although neither taken separately entails E. Scientists are willing to believe E if they believe in T_1 and T_2. But if the scientists are merely instrumentalists of T_1 and T_2 by believing that they are merely empirically adequate, there is no reason to believe in E. You need to be a realist about T_1 and T_2 in order to believe E. If you believe only in an empirical adequacy

[17] Again, in Friedman's example, we can say that the theory of gases is less confirmed as itself than when it is mapped on the theoretical structure containing Newtonian laws of mechanics.

of theories without any theoretical support, there is no guarantee that the conjunction of such two theories is empirically adequate, unless there exists a theoretical structure that unifies them.

Similar to Putnam, Friedman holds that a literal theory evolves by conjunction, whereas a non-literal theory does not. In the non-literal construction, there is no real molecular world out there; the only reality is that of gases and the phenomenological measures, i.e. the phenomenological world $\mathcal{B}$. But for a realist, theories evolve by conjunctively adding something to $\mathcal{A}$ and *not* by stipulating in time different maps from different theories $\mathcal{A}_1$ and $\mathcal{A}_2$ onto the same $\mathcal{B}$. A valid argument in the *literal* interpretation evolves by conjunction in the sense that if a model $\mathcal{A}$ that postulates at t_1 some relations R_1 is part of a class of models Δ_1 and then it evolves by postulating some other relations R_1 and now belongs to a class Δ_2 at t_2, then the model that postulates R_1 and R_2 belongs to the intersection of Δ_1 and Δ_2:

$$\begin{array}{c} \langle B, R_1 \rangle \supseteq \mathcal{A} \text{ and } \mathcal{A} \in \Delta_1 \\ \langle B, R_2 \rangle \supseteq \mathcal{A} \text{ and } \mathcal{A} \in \Delta_2 \\ \hline \therefore \langle B, R_1, R_2 \rangle \supseteq \mathcal{A} \text{ and } \mathcal{A} \in \Delta_1 \cap \Delta_2 \end{array} \tag{1}$$

In the representational interpretation, at t_1 the phenomenological structure $\langle B, R_1 \rangle$ is (only) mapped onto $\mathcal{A}$ by a map ϕ and mapped at t_2 onto a different $\mathcal{A}'$ by a different map ψ:

$$\begin{array}{c} \exists \mathcal{A}, \exists \phi : \langle B, R_1 \rangle \rightarrow \mathcal{A} \text{ and } \mathcal{A} \in \Delta_1 \\ \exists \mathcal{A}', \exists \psi : \langle B, R_2 \rangle \rightarrow \mathcal{A}' \text{ and } \mathcal{A}' \in \Delta_2 \\ \hline \therefore \exists \mathcal{A}'', \exists \chi : \langle B, R_1 R_2 \rangle \rightarrow \mathcal{A}'' \text{ and } \mathcal{A}'' \in \Delta_1 \cap \Delta_2 \end{array} \tag{2}$$

The argument (2) is invalid, whereas (1) is valid because we have different terms in the two premises. Here is the major difference between literal and non-literal interpretations: "the conjunction of two reductions implies a single joint reduction; the conjunction of two representations does not in general imply a single joint representation." (Friedman 1983 246). After a period of time, a new observational prediction can confirm the conjunction of two theories without following from them separately. If one sticks to the representational description of the theory, the evidence A is not boosted repeatedly. If we adopt a literal scheme for this, we have:

$$\frac{\begin{array}{l}\text{If at } t_1 \text{ A is confirmed}\\ \text{If at } t_2 \text{ B is confirmed}\end{array}}{\therefore \text{ At } t_3 > t_1,\ t_2 \text{ both A and B are confirmed}}$$

whereas in a representational model we have:

$$\frac{\begin{array}{l}\text{If at } t_1 \text{ A is confirmed}\\ \text{If at } t_2 \text{ B is confirmed}\end{array}}{\therefore \text{ At } t_3 > t_1,\ t_2\ \exists\chi \text{ such that } \chi(A \& B) \text{ is confirmed and } \sim\left(\exists\phi(A) \& \exists\psi(B)\right)}$$

As such, $\chi(A\&B)$ is not formulated till t_3, so it is not subjected as the same kind of test as A and B at t_3.

If at an initial moment of time a set of observational data $\mathcal{B}_1$ are successfully derived from a theoretical structure $\mathcal{A}$, then at this stage, the literal interpretation will benefit in general of a smaller degree of confirmation than the representational interpretation. If at a later moment of time the same structure is used to derive the properties of a different observational structure $\mathcal{B}_2$, then, according to Friedman, only the literal interpretation of $\mathcal{A}$ will receive a boost in confirmation. In time, after several successful derivations of different observational structures $\mathcal{B}_1, \mathcal{B}_2 \ldots \mathcal{B}_n$ from the same theoretical structure $\mathcal{A}$, the literal interpretation of $\mathcal{A}$ is preferable.

Therefore, in the representational model no hypothesis receives repeated confirmation in time. Friedman's conjecture is that our theories evolve by conjunction, so the representational model is incomplete and is not desirable.

Another crucial point is that unification is a *relative* feature of the theory: a part of the theoretical structure of a theory $\mathcal{A}_1$ is unificatory only in the context of a second theory $\mathcal{A}_2$. Friedman's example is very challenging: absolute rest does not unify in the context of Newton's theory of gravity, but it does unify in the context of electrodynamics. If ϕ has a necessary role in many inferences, then it has to be taken literally and it has a rightful place in the physical reality $\mathcal{A}$.

The historical story to be told by Friedman is like this: the starting point is a structure $\mathcal{A}$ which a representational structure is added to. The elements of the representational structure are not taken literally, they are simply mathematical structures such as coordinates, units, etc. Some elements of $\mathcal{A}$ are assigned to the real world, and some to its mathematical representation: in some spacetime theories, $\mathcal{A}$ is $\mathbb{R}^4$. Finally, a particular piece of structure ϕ postulated by the theory can be considered unificatory in the context of ψ if it can facilitate the following inference:

$$\begin{array}{c} \exists \phi \in \Phi : \mathcal{A} \to \mathbb{R}^4 \\ \exists \psi \in \Psi : \mathcal{A} \to \mathbb{R}^4 \\ \hline \exists \chi \in \Phi \cap \Psi : \mathcal{A} \to \mathbb{R}^4 \end{array} \tag{3}$$

If this is valid without ϕ, then it has no unificatory power and it can be dropped from $\mathcal{A}$. If it is unificatory, then it has to be kept and then interpreted literally. Otherwise the theory is less confirmed.

There are several criticisms of Friedman's argument worth mentioning. First, Morrison suspects that Friedman needs realism to define unification and then he constructs the argument for realism based on unification; in other words, Friedman's argument seems circular. This suspicion, echoed by A. Kukla, is similar basis to the argument used by A. Fine and L. Laudan against the original no-miracle argument (Morrison 2000 37; Kukla 1995 235). I discuss Morrison's attack on Friedman in Chapter 5, where I will deal with several details of her account.

Second, for Morrison, van Frasseen and Kukla, the very idea that science evolves by conjunction and that the only possible interpretation of theoretical terms is the literal one is very problematic given the history of science and the way science evolved: "when we conjoin theories we rarely, if ever, do so strictly on the basis of logical principles" (Morrison 2000 42); cf. (Van Fraassen 1980, 235). Based on the philosophy of science of the last decades, these authors conclude that Friedman's model is too simplistic and too abstract to describe the complex evolution of scientific theories in time. Third, for Morrison, Friedman had collapsed the difference between epistemological and semantic realism and such confusion can have several undesirable consequences. Another question is whether one can separate the phenomenological structure from the theoretical one as easy and clear as Friedman wants to. What if there is more than one way to identify the two structures, even literally? If the observational structure is already theory-laden, how do we separate $\mathcal{A}$ and $\mathcal{B}$?

My own criticism against Friedman is twofold. First, it seems that Friedman omits a major fact about science. Science intentionally mis-describes the world in order to represent it by the means of idealizations and simplifications and this procedure is present within unificatory attempts, as well (Batterman 2005). In many cases this is the only way to access reality. How important are fictional entities in the economy of unification? Friedman seems to give them no importance at all. In general, Friedman's account does not capture the idea of surplus structure that is introduced by the formalism and entertained temporarily by theories. Second, there is a clear sense in which conjunction is not what unification is supposed to be. Some trivial or spurious unification are mere conjunction, but Friedman's centrality of conjunction in his account of unification seems a deadlock. In my case study I will try to make a distinction between unifications based on conjunction and other unification. Think of **EM** theory: at a first sight, it has nothing to do with conjunction because a Electricity & Magnetism theory, literally interpreted, does not account for the host of phenomena Maxwell's theory was able to account for. It seems that Friedman's account misses completely the idea of a coupling element between T_1 and T_2 which is never present in their conjunction. Moreover, while conjunction is truth-conducive, this is not in general true for unification. One can see why logical truth and logical conjunction as used in Friedman's account simply do not fit the way unification is present in science.

In short, here are Friedman's (1983) claims discussed above:

[1] *By the means of conjunction, theoretical unification entails better confirmation of theories.*

[2] *Increased empirical confirmation increases the likelihood that the theory is true.*

[3] *Hence, unification entails realism.*

As a prelude to Part III, I mention here that my case study will disconfirm Friedman's account of unification by conjunction. Unlike the ether or Newtonian absolute spacetime, in the case of Kaluza-Klein theory, nobody knew 70 years ago, and we still do not know—whether extra dimensions have to be interpreted literally as existing in the physical space or as internal degrees of freedom that we could attribute to space-time. In this dissertation I do not assume that Kaluza's or Klein's theories—at least at the historical stages exposed here—proceed by conjunction and I do not assume that unification is truth conducive. I want to emphasize the role of corrective unification and the differences between a conjunctive phase of a theory and its later, more creative and innovative phase. There is also a question of whether new Kaluza-Klein theories as formulated in the 1980s can be used as an argument for the existence of extra-spatial dimensions. Most probably, Kaluza-Klein theories, in almost all incarnations are not literally true but they tell us important things about the dimensionality of the real world which is not a direct observational fact—at least not yet: some think that the signature of extra spatial dimensions could become available by the experiments at LHC.[18] This is an open discussion and the jury is still out until stronger empirical evidence has been collected possibly at the LHC. What is important is that Kaluza-Klein theories are not true in a *literal* sense *à la* Friedman and they did not evolve by conjunction either. The coupling element, not discussed in Friedman can actually materialize in the so-called Kaluza-Klein particles (Randall 2005, 499). If there is an inference from unification to realism it should be based on mere conjunction. What I am interested in discussing Friedman's argument for realism based on unification, is whether unificatory theories, interpreted literally, are confirmed or not by conjunctions, and whether those interpreted phenomenologically (or non-literally) are not or are confirmed as conjunctions.

[18] The theory is discussed in (Randall 1999).

1.6. The quest after unification in science

The enthusiasm for unification share by some philosophers is shared by many scientists. The aforementioned philosophers' enthusiasm for unification has a counterpart especially in the way the domain of high energy physics was regarded in the last four decades. History of physics in general can be read as a partial history of successive unifications, but the story of the 20th century and of the present decade can hardly be told without stressing the desire for unification: Einstein's unified field theory, various Grand Unified Theories (**GUT**), Supersymmery, Superstring theory, Canonical Quantum Gravity, and many more. Other theories which apply to a wide range of phenomena such as statistical mechanics or quantum mechanics also have unification as one of their motivations.

Within theoretical physics itself, unification can be understood in several ways: some unificatory programs were designed to unify fundamental fields, some were designed to unify matter with fields, and yet others were premised on even stronger assumptions and endeavored to unify gravity with all the other known fields (Weingard 1991, nd). Cosmology aims to provide a unified picture of the universe from the earliest stages to its distant future. The "consensus model" in cosmology is premised on the idea that one and the same physics with a set of constants has governed the evolution of the universe since its beginning (the standard name of this consensus model is "Lambda + cold dark matter" (**LCDM**), although alternatives to it are also attractive, for example P. Steinhardt's model is premised on the idea that different physics had acted at different epochs during the evolution of the universe (Steinhardt and Turok 2007, 284). Supersymmetry aims to relate the properties of bosons to the properties of fermions and, moreover, to align the property of a particle to the property of its super-partner. Here we see a unificatory theory which comes with its own baggage of problems: it postulates the existence of particles which are not observable yet.

The foremost unification of all unifications is the synthesis of quantum mechanics with the theory of general relativity, or put it metaphorically, to unify “the discrete” and “continuous” aspects of reality. Several flavors of quantum gravity programs and String Theory compete in achieving this “unification of all unifications” and of course, needless to say, each comes with its own “fine prints”. Criticisms against String Theory are already popular (Smolin 2006; Woit 2006, 291).[19] Besides these, there are speculations that once formulated, the mysterious “M-theory”—circulated first only among the String Theory community, but now known to everybody as the most unificatory theory of all—will be able to unify mathematics and physics.[20] The majority of philosophers are agnostic in respect of String Theory: they hypothesize it is too early to analyze a highly incomplete theory “not ready for certain kinds of foundational studies” (Weingard 1988).

Even given possible troubles in the paradise of unification, the majority of physicists would endorse an architectonic representation of known interactions that can be read as a progressive history toward unification. After confirming the existence of four fundamental physical forces—all the other forces being merely apparent or derivative from these: electromagnetism, (being already unified), gravity, the strong nuclear force, and the weak nuclear force, in the first half of the 20th century and developing accurate theories of these forces for each of them: “the aim of physics is now to produce theories which unify

[19] A critical overview of Loop Quantum Gravity and its claim of unification is found in (Nicolai, Peeters, and Zamaklar 2005).

[20] The “M-theory” was popularized by E. Witten in 1995-1996. It was originally thought that this 11-D theory is more fundamental and unifies all known string theories. “M” could stand for: “mystery”, “magical”, “matrix” “membrane”, or even for “mother”.

these forces, which show, ultimately, that there is at base only one fundamental force in the universe, which has come to display itself as if it were many different forces" (Maudlin 1996 129).

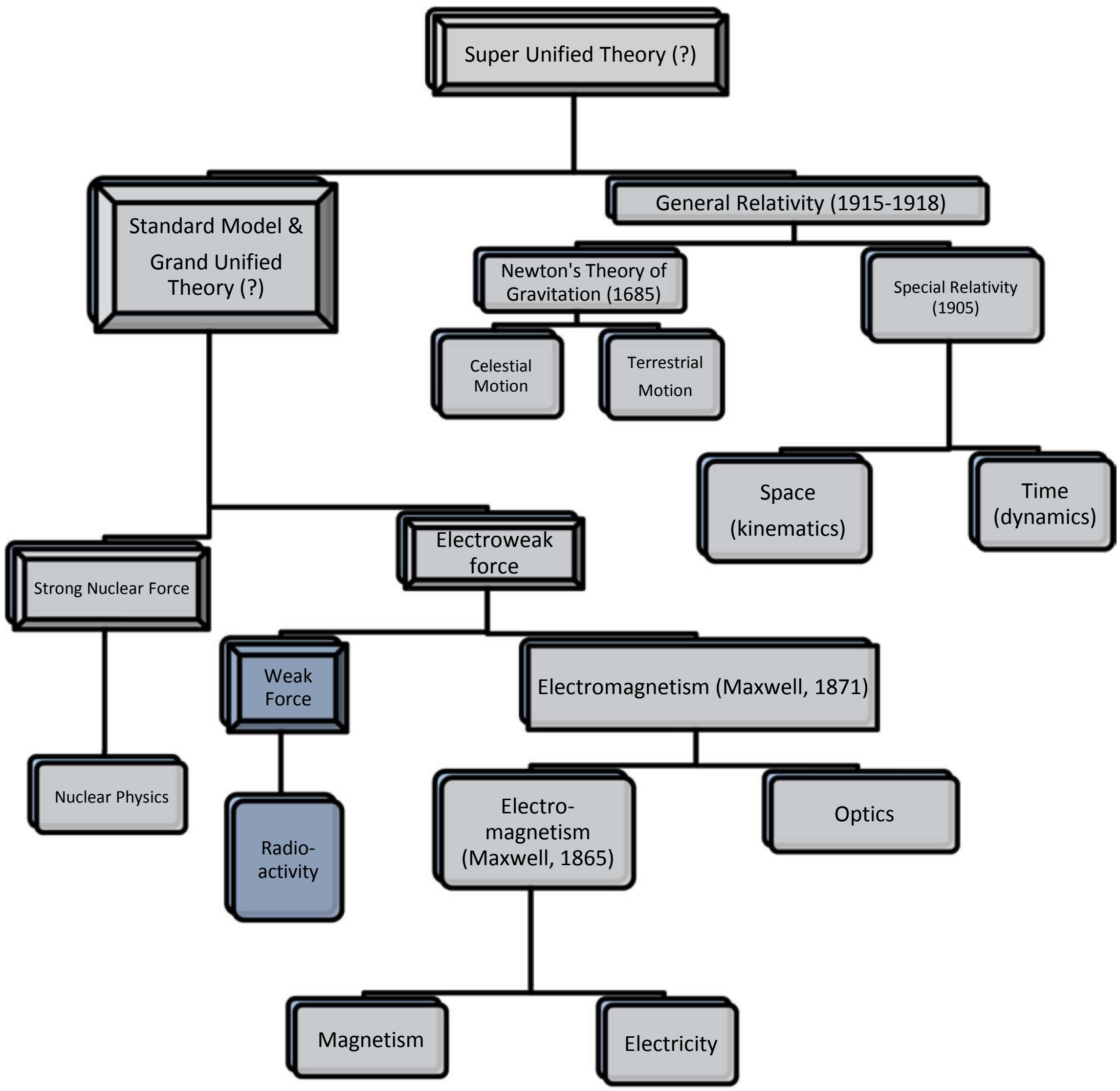

Figure 1 (elevated rectangles represent quantum theories)

One can see that successfully achieved or only dreamt of, unification is avidly sought in theoretical physics. S. Glashow suggested that in the 1950s, after the huge success of quantum field theories, physics was "patchy":

> the study of elementary particles was like a patchwork quilt. Electrodynamics, weak interactions, and *strong* interactions were clearly separate disciplines, separately taught and separately studied. There *was* no coherent theory that described them all. Developments such as the observation of parity-violation, the successes of quantum electrodynamics, the discovery of hadron resonances and the appearance of strangeness were well-defined *parts of* the picture, but they could not be easily fitted together (Glashow 1980 539).

Praising the standard model of elementary particles, Glashow claimed that in the mid-1970s, it had already constituted a complete and apparently correct theory, postulating a small number of fundamental masses and coupling constants:

> The theory we now have is an integral work of art: patchwork quilt has become a tapestry. […] Tapestries are made by many artisans working together. The contributions of separate workers cannot be discerned in the completed work, and the loose and false threads have been covered over. So it is in our picture of particle physics. Part of the picture is the unification of weak and electromagnetic interactions and the prediction of neutral currents, now being celebrated by the award of the Nobel Prize. Another part concerns the reasoned evolution of the quark hypothesis from mere whimsy to established dogma. Yet another is the development of quantum chromodynamics into a plausible, power and predictive theory of strong interactions. All is woven together in the tapestry; one part makes little sense without the other (539).

Notwithstanding some notable exceptions, this was the received view among physicists and still remains the dominant consensus. Nowadays, Glashow's optimism is reflected by the attitude of some physicists—except towards gravity…

Weinberg suggested that the physicists' job is to unify our representation of the world: "Our job as physicists is to see things simply, to understand a great many complicated phenomena in a unified way, in terms of a few simple principles" (Weinberg 1980 515) More recently, in an interview on B. Greene's book, *The Elegant Universe*, Weinberg was asked why the claim of unification was so important (in the context of String Theory). His answer reveals maybe one of the most optimistic stances toward unification:

> Unification is where it's at. The whole aim of fundamental physics is to see more and more of the world's phenomena in terms of fewer and fewer and simpler and simpler principles. And the way you do this is not by having one book on electromagnetism, and another book on the weak interactions, and so on, but to have just one book on all the forces of nature. A simpler description—that's what we're aiming at (Weinberg 2003; referring to Greene 1999).

This enthusiasm toward unification is less frequent among biologists or chemists, where the fragmentation in parochial fields is perhaps more blatant than in theoretical sciences. In some fields such as molecular genetics or oxidative metabolism, there is huge fragmentation—in part because the same processes do not operate in all orders of life or in the same manner. Nonetheless, some think that biology has reached the level at which the process of fragmentation will be replaced with a steady process of consolidation. The most enthusiast scientists see consolidation as a sign of unification:

> Scientific progress is based ultimately on unification rather than fragmentation of knowledge. At the threshold of what is widely regarded as the century of biology, the life sciences are undergoing a profound transformation. They have long existed as a collection of narrow, even parochial, disciplines with well-defined territories. Now they are undergoing consolidation, forming two major domains: one extending from the molecule to the organism, the other bringing together population biology, biodiversity studies, and ecology. Kept separate, these domains, no matter how fruitful, cannot hope to deliver on the full promise of modern biology. They cannot lead to an appreciation of life in its full complexity, from the molecule to the biosphere, nor to the generation of maximal benefits to medicine, industry, agriculture, or conservation biology. (Kafatos and Eisner 2004 1257).

Kafatos and Eisner suggest an interesting idea: unification is present preponderantly at the mature stage of a theory. Their view is totally opposed to the dissenters who say that unification is a misleading chimera: it belongs to the undeveloped stage of a science. Some sciences look unified, but as they progress they become more and more oriented towards details and particular phenomena. For the dissenter, scientific progress increases fragmentation and disunity and disunified sciences better describe the world in details which are not accessible to the more unified theory.

I think the whole discussion about unification or fragmentation in general is a question of degree and nuance. A totally dis-unified science and a totally unified science are both impossible. I also see how easily both positions can be distorted and exaggerated *ad nauseam*. Some of the above claims are trivially true and they do not tell us much about how science works. If one admits that there is a common scientific method, a set of common standards and norms, and maybe a scientific language, then *voilà*: there is unification. Strong, ironclad, dissenters would disagree on all of the above. On the other hand, I claim that a completely dis-unified, totally fragmented science, is impossible. Any science displays a certain degree of unity, at least at the level of reasoning, language or standards. Creating a totally isolated field is not a scientific enterprise, but maybe something close to an ideology or a dogma. Because of theory-ladeness, confirmation and measurement imply a specific theoretical framework. There is always interconnectedness and a web of reciprocal dependencies between theories and this can be interpreted as a form of unity of science. A weak form of holism is necessary to any discussion on unity and unification. The use of scientific language, the mathematical method, the way we visualize and represent the world, are all common features of many scientific theories.

Scientists and some philosophers are indeed fond of this sense of unity. Kitcher endorses this view in a recent debate with the advocates of dis-unity (N. Cartwright, J. Dupré, I. Hacking, etc.):

> Workers in any field know how their projects relate to those of their colleagues, and, typically, see their own endeavors as contributing to very general questions about nature. The collective research is structured by very general schemes of explanation that unify the phenomena of the field (Kitcher 1999 338).

In respect of unity, I am worried that it can be easily trivialized. Needless to say, almost all aforementioned claims about unity are controversial to a lesser or a greater extent as not everybody buys an overly optimistic view. I do not want to suggest that anything two theories have in common makes for unity. The main question is how much mileage one gets from this intuition of unity in concrete cases of mature scientific theories and how local, or—on the contrary—how global this feature is. As I employ it here, unification is definitively more precise a concept than unity. I suggest that having something in common such as: method, mathematical representation, standards are only necessary conditions to achieve unification. The unificatory theory is not built from the two previous theories on the basis of vague similarities or analogies or mere on the basis of a common knowledge. There should be another ground for unification than the vague concept of unity.

Chapter 2. Arguments against unification

Unification improves upon two existing theories in respect of explanation, ad-hocness, simplicity, etc. If the strong positions such as Friedman's or Whewell's are correct, then we can even start to believe more in unificatory theories than in the dis-unified theories. There are some dissenters from the favorable assessment of unification, both in philosophy and in physics.

2.1. Philosophical skepticism against unification

The deflationism view: unification as accidental. Firstly, on a moderate reading, a modest dissenter can deny the relevance of unification in general. Roughly put, although she admits that there are occurrences of unification in science, a more unificatory theory is nothing special. It is an accidental feature of a theory and by no means a virtue, empirical or non-empirical. As a contingent feature of some theories, it can accompany or not their explanatory power, their realism, their simplicity, etc. It is not that we seek it and it does not occur at a mature stage of a theory—it can be otherwise. It does not mean that unification is *never* packaged with other virtues such as explanation or simplicity; such cases can exist and they are indeed felicitous. But in general a unificatory theory does not explain, neither predicts, nor is better confirmed than other theories.[21] The modest dissenter may actually admit that some unificatory theories can be *only accidentally* explanatory, simpler, even closer to truth than the dis-unified theories.

[21] Similar claims were recently inferred from Myrvold's account of unification by Schupbach. (Schupbach 2005).

The acausality of unification. For the so-called "ontic approach" to explanation, best illustrated by W. Salmon's "causal-mechanical" approach, to explain is to reveal the causal structure (*nexus*) of the world: we explain events by showing how they fit into the causal nexus" (Salmon 1984 276). Later on, sophisticated tools based on probability were developed to clarify this basic idea.[22] Discovering causal processes linking individual events and patterns, too, is the business of physical science. In general, science explains by revealing the causal mechanisms hidden under the phenomena. So is there a place for unification is this causation+explanation party of two?

Kitcher's early views (from the 1970s to the mid-1990s) were at odds with this long-lasting philosophical tradition. For him, unification is constitutive to explanation and causation comes into play later.[23] Salmon called this approach the "epistemic approach" (in contrast to his own "ontic approach"). For the dissenters, Kitcher distorted or omitted scientific explanation because his model sins precisely where the deductive-nomological model (**D-N**) found its own death.[24] The example that has posed a serious challenge for the **D-N** model was the explanation of the length of shadow of a flagpole in sunlight. Indeed the unificationist can account for the fact that the height of the flagpole, together with the position of the Sun, some laws of optics, geometry, the theory of the propagation of light in air, etc. explain the length of its shadow because the shadow is caused by the interaction between sunlight, flagpole, and their relative positions to the ground. But the unificationists had a hard time in eliminating the explanation that runs the other way around: the

[22] Initiated by P. Suppes, and later developed by C. Glymour, J. Woodward, R. Spirtes, etc. (Glymour 1980; Spirtes 1991, 1-36; Woodward 1984, 231-262) I discuss extensively J. Woodward's recent critique of Kitcher (Woodward 2003 ch. 8) in Section 4.2.

[23] This is an extensive philosophical tradition, attributed to Aristotle, Bacon, and especially to Kant.

[24] I discuss in details a criticisms to Kitcher by Woodward in Section 4.2 and the D-N model in Section 3.1.

length of the shadow can explain in fact the height of the flagpole. The epistemic approach and the **D-N** model get the forward causation right, but they take the backward explanation as genuine, too. The epistemic approach could not represent, in a desirable manner, the causal relations and especially the inherent causal asymmetry. Similar worries against unification can be found in (Jones 1995a; Barnes 1992; Schurz and Lambert 1994; Hilpinen 1980; Schurz 1999).

Another quick word on Salmon: it is wrong to think that he dismissed unification altogether; rather he viewed unification as a *consequence* of causal explanation. The ontic approach has as much right as the epistemic approach (Kitcher's unificationism) in taking unification of phenomena as a central aspect of our way of representing the world.[25]

In a similar vein, Cartwright contends that regulatory laws are not the right framework for understanding the unificatory power of the principles of Newtonian physics. She urges us to render unification in claims about capacities that can be assembled and reassembled in different nomological machines (Cartwright 1999b 52). Similar to Salmon, and in some extent similar to what Morrison hints at, Cartwright suggests that there is unification in science, but it is a consequence of an underlying structure (causal, dispositional, of capacities etc.). On this dis-unified patchwork picture, science does what it can relative to certain strategies, phenomena and theories, but it does not need to bring all these different "patches" together into a single consistent picture. Cartwright, one of the main advocates of this approach, does not deny that principles of physics are unificatory, but she renders this in a different language, that of capacities and natures instead of laws:

[25] See Chapter 4.1 for a discussion of Salmon in the context of Kitcher's approach.

> The example of the planetary motions [see Kepler's and Newton's theories] is important for me since it has been used by philosophers and physicists alike in support of the view that holds more 'basic' regularities as first and fundamental in accounting for observed regularities (i.e., in explanation, it is laws 'all the way down'). This view emphasizes the unifying power of the appeal to Newton's laws with respect to Kepler's. I do not deny the unifying power of the principles of physics. But I do deny that these principles can generally be reconstructed as regularity laws. If one wants to see their unifying power, they are far better rendered as claims about capacities, capacities that can be assembled and reassembled in different nomological machines, unending in their variety, to give rise to different laws (Sklar 2003 52).

I subsume her views, and maybe more remotely Salmon's and Morrison's, under the slogan: "unification, unification, so what? Causation (capacities or whatever) are more important."

Although there is a lot more to be said on the topic, intuitively at least, causal explanation, a.k.a. the "ontic approach" and the unificatory approach, more or less epistemic, can significantly diverge (Gijsberg 2007). The way in which Morrison critically examined Friedman's and Kitcher's model reflects the causal stance toward explanation: her dissatisfaction with different instances of unification can be translated in the worry that unification does not properly render the causal mechanism, or does not place the emphasis on the causal mechanism.[26] One can have the feeling that at this stage the philosophers who take unification as fundamental and not as a derivative notion are a minority worth defending. A reason to take their side is my skepticism that causal explanations are the best or the most relevant explanations in any area of science. Contra Morrison, I think unification operates even where causation does not play a major role or where we do not have or cannot expect to have a full causal story.

---

[26] See Chapter 5 on Morrison.

The anomalous unification. For those who believe that unification is too expensive, there is a drawback of unificatory theories. Even if unification is possible and in many cases it works, some authors remind us that there is a price to pay: unification brings in complexity, ugliness, incoherence, etc. For example, in a recent paper about plurality in **QM**, M. Dickson thinks that monism (a term similar to what I coined here "unification") brings in "anomalism": although "diverse theories, interpretations, or methodologies can be consistently conjoined into a single theory or methodology", this monistic, unified theory is not systematic and it is not formulated in a lawlike way. It has the features of a mere conjunction "rather than any sort of union":

> The view is anomalous (*a-nomos*, from Greek) because the diverse elements of the single truth, interpretation, or methodology can only be conjoined to form the one, and cannot be united onto a single theory from which the diverse elements are derived in any lawlike way. [...] no one of the diverse theories, interpretations, or methodologies can be reduced to another in a lawlike or systematic way. Their terms, for example, may not be interdefinable (Dickson 2006 43).

If lawlikeness defines a scientific theory, by unification it may be lost. It is important to see whether the price to pay includes the unificatory theory being lawless or anomalous (or, better, less nomic that T_1 and T_2). What is too restrictive in Dickson's view is the reduction he expects from unification. Yet unification may be worth the value of the goal of explanation.

The physics-chauvinism of unification. Another pool of philosophers would doubt that we can achieve unification in areas other than theoretical physics. Except in some isolated cases, special sciences progress by creating new disciplines more or less remote from their foundation. Since Popper, boastful unificatory programs in social sciences are in general viewed with an unfriendly eye. One can imagine a pseudo-theory or a pseudo-science that has all the theoretical virtues you can dream of, but which is blatantly false. Standard examples, some analyzed by Popper (and later by Lakatos), include Marxism, psychoanalysis, astrology, monetarism, etc. which are truly unified theories, with indubitable explanatory powers (one can say that they are superexplanatory or overexplanatory theories) some of them are simple theories and/or elegant or more attractive than their components (Popper 1962 34-5). But all of them score very badly in respect of ad-hocness, empirical adequacy, novel predictions and of course, they are on almost all accounts, false. According to these dissenters, unification in theoretical physics may or may not work, but outside this area it is not prevailing. One can suggest that the worst cases of spurious, trivial and blatantly bad unifications can be found in the special sciences. Bad unifications abound in physics, too: one of the worst unificatory hypotheses in physics is maybe the idea that ether *unifies* all interactions. The same can be said regarding phlogiston: it unified a large body of empirical knowledge, but the oxygen—the real element that explains—explains in fact a smaller range than phlogiston. Both ether and phlogiston were unificatory and explanatory, but they were abandoned for several reasons.[27] Abandoning the ether came with a higher price for causal explanation: its replacement was something less causal and less concrete: the empty space in which electromagnetic waves can exist replaced ether.

A standard strategy would be to take unificatory techniques, methodologies and strategies from the most successful sciences and import them unconditionally in other sciences (mainly in special sciences). Cartwright expresses a worry against the universality

[27] It is a different thing to talk about vacuum as *unifying* several forces. Here it is understood more or less as a mathematical condition and not as an entity like ether (Aitchison 1991).

of the exact sciences: “[…] I am worried that an ill-supported belief in the universality of our favorite exact science can lead us to adopt bad methodologies for carrying out this [the universality] aim.” (Cartwright 1999a 333). I raise the same worry in respect of unification: trying to impose unification found in theoretical physics as a standard upon special sciences may harm more than help. Why is there all this talk about scientific unification when the only unification is found in theoretical physics (if even there)? There are bad unifications and good unifications in physics and it is not clear whether unification is present in other sciences: there may be bad unifications and good unification is special sciences, too. Even if the status or even the existence of unification in special or non-fundamental sciences is unclear, to evaluate any such attempts, philosophers of science should better know how unification in physics was achieved or perhaps how it failed. Exemplar and failed unifications are to be found predominately in physics. In my view, the analysis of unification in physics should be descriptive and not normative or prescriptive in respect to special sciences.

Unification without laws and without deduction. The exemplar cases of unification are usually formulated as mathematical equations that gain generality and universality and are more fundamental than previous laws (or equations). There are emerging pictures of science whereby the desideratum of science is not to provide a fundamental law from which everything can be derived. The mechanistic view endeavors to explain through a different route than laws of nature. Mechanistic explanations work by identifying component parts and operations of a mechanism and figuring out how they are organized so as to generate the phenomenon. W. Bechtel and A. Abrahamsen claim that generalization is possible within the mechanistic view, without the presence of laws. Some explanations of a class of specific phenomena can be generalized to similar, related phenomena. Thus, discovering mechanistic explanations is possible (Bechtel and Abrahamsen 2005 432). If explanation is possible without laws, is there unification without laws of nature? One quick answer would be that unification entails the identification of mechanisms. S. Glennan argues that the facile way of deriving unification from the unity of fundamental mechanisms is wrong, "although explanatory unification afforded by the mechanistic approach derives not only from the commonality of fundamental laws but from the existence of mechanisms that have a common higher-level structure even if they differ in microstructure" (Glennan 2002 S352). Again the argument can be read as endorsing the idea that identification and identity is not enough to achieve unification and moreover unification can be fulfilled by different means. Then there is unification in sciences where laws do not play the crucial role in explanation and prediction. On the contrary to what Dickson's suggested, being less nomical or even anomalous does not necessarily demote unification (see p. 14).

This discussion is partially related to my case study because both Kaluza and Klein heavily used laws of nature in deriving their theories. Causation was not present in the original formulation of their theories although it can be added later on. I defuse the role of causal explanation or causal unification in my case study by arguing that other reasons to unify are at stake here. The plan is to discuss Kaluza's and Klein's projects as types of geometrization. The explanation at work in Kaluza and Klein is not explicitly causal, but geometrical, as I argue in Part III. There is also a way on interpreting Kaluza-Klein theory close to the spirit of mechanistic explanations.[28] It is also true that there is a mechanism of the compactification of extra dimensions which is basically a causal story trying to answer questions like: “why is the fifth dimension compactified?” “what is the cause of compactification?”. A further claim to be investigated is whether *mechanistic* explanations and implicitly mechanistic unifications could be found in theoretical physics, too. The unification rooted in mathematics can coexist with mechanistic unification in theoretical physics. It is an open question whether there is a mechanistic unification in later incarnations of the Kaluza-Klein theories.

Surplus mathematics and unification. For some other dissenters, unification is sometimes disguised in theoretical structures, and theoretical structures are sometimes problematic because we do not know whether they are real or only a contrivance of our representation of the world: for example think of the gauge potentials. Because not all theoretical structures are related to reality, a nominalist asks questions like: What is a theoretical structure good for? Do we really need this and that parameter? Do they explain or they are only crutches to our lame representations of the world? Similar questions were asked again and again in physics in the contexts of gauge theories, Feynman currents, quantum mechanics, algebraic quantum field theory, spontaneous symmetry breaking and for concrete entities such as the graviton, gravitational waves, etc.[29]

We know that theoretical structures are usually larger than phenomenological structures—they are abstract, unobservable, idealized, etc. By surplus structure I mean

---

[28] I refer here to a mechanism of “spontaneous symmetry breaking” related to a specific type of Kaluza-Klein theory.

[29] For example, a detector with the mass of a massive planet and with 100% efficiency, placed very close to a neutron star, would detect one graviton every 10 years, even under the most favorable conditions (Rothman and Boughn 2006, 1801-1825).

parts of a mathematical formalism that do not correspond to anything in reality, but which are useful for computations or come out as results of computations.[30] For Readhead and Teller, less surplus structure in the formalism of one theory should be a reason for choosing that theory over another loaded with surplus structure. Contra Redhead and Teller, I claim that even if the surplus structure is not observable, it may play a methodological role.

Some debunkers of unification could say that unification is *always* couched in terms of surplus structure or, worse, of *bad* mathematical structures that may look elegant or even useful. Unification is theoretical, or more precisely, mathematical trickery. It may be computationally and operationally useful at some stages of a theory's development, but unrelated to anything in the real world. It covers some computational or structural failures of our theories or of our models of the world, whereas the world in itself is totally dis-unified. We need to extricate ourselves as soon as possible from this "folk unification" which is a mere illusion. For positivists, any kind of theoretical explanations resulting from any kind of theoretical structure should always be seen as a convenient device for generating empirical consequences. Besides this role, theoretical structures have no reality whatsoever; they are mere auxiliary tools for representing the observable phenomena. In this respect, one can see why parts of theoretical physics, especially String Theory, were

[30] A short description of what "theoretical structure" is can be found in (Friedman 1981 1). In the received view, surplus structures are unphysical and they have to be eliminated when we have no observational data of them (Redhead and Teller 1992; Redhead 1975, 77-112). In other cases, what had been thought as surplus in P. M. Dirac's discovery of the positron, i.e. the negative energy solutions, would turn out as corresponding to real particle (Dirac used the words "irrelevancies"). We should entertain surplus structure in science, at least for a while. The status of gauge potentials as a surplus structure is still a hot topic in the philosophy of physics.

dubbed “science fiction in mathematical form” being on the verge of religion and magic (Horgan 1996).[31] I will revisit this issue in the next section.

This dissenter’s view on unification originates in the skeptical attitude toward theoretical structures in general because, on several accounts, unification depends exclusively on such theoretical structures. If unification is mainly an attribute of our theories or, more precisely, of our representation of the world, how do we relate it to the unity of nature, if any? Take for example electricity and magnetism. They are either one thing or not, independently of our wishes and desires. Nevertheless, electromagnetism, truly one of the exemplar unificatory theories at hand, says that they are in a sense one thing (or better, manifestations of one lower-level entity). Another bolder claim one can find in today’s elementary particle physics, for example the so-called “Minimal Supersymmetric Standard Model”, refers to bosons and fermions: are they really one thing? They either are or are not, regardless of whether we can find a unified representational framework in which they are one.[32] If unification operates at the level of our representation of the world, why believe that it says something deeper about science or about the world in general? There is an immediate answer to this question. Our beliefs about the world are justified via the scientific representation and for a large class of phenomena mathematical representations in science score better than alternative, non-mathematical representations. If some unification turn out to be empty or spurious in this sense, that does not mean that any unification whatsoever are empty or spurious.

---

[31] As some bloggers on the Internet suggest, String theorists will never host a “Reality Show”.

[32] A piecemeal solution is to accept that they are the same *at very high energy* to which we do not have yet access. If so, the gauge couplings of the three gauge groups (electromagnetic, weak and strong) unify in reality, but for the moment we can only represent that reality without confirming it.

Fragmentation of science. The dis-unity movement does not attack directly the idea of unification (B), but the unity of science (A) which is more general and, as I take it, not directly related to (B) (see Section 1.2). According to a more or less radical form of pluralism, there are always an indefinite number of ways of individuating and classifying the objects on the world (Dupré 1993, 308). A more modest pluralism stays away from metaphysical claims about how the world is and focuses on the plurality of our scientific theories. Even if the world is “unitary”, some parts of it are some complicated that we need to use different theories, languages and representations to gain any knowledge of it. There is no single representational idiom and no way to differentiate the best one (Kellert, Longino, and Waters 2006 xii). The complete representation of phenomena requires multiple approaches.

If you buy Cartwright’s view that we actually live in a “dappled world”—covered with patches of theory cut up in countless ways—and you keep your science dis-unified, another benefit is that independent parts of science can assist in independent, objective testing of each other’s claims. Some suggest that this avoids theory-laden observations and bolsters objectivity (Kosso 1989; Hacking 1983).[33] Anextreme position is to hold that the ultimate **TOE** would need no recourse to experiment. A **TOE** would not be less falsifiable than our current theories. Everything would be defined in terms of other elements of the same theory and experiment will merely serve to define a convention of scaling between parts of the **TOE** (Davies and Brown 1988 7).

All these suggest that the main target is in fact (A) and not (B), as well as unity of science and not unification. So pluralism and unification can coexist, can’t they? Some-

[33] But this view was challenged by (Roush 2004).

times and perhaps indirectly, unification *per se* is targeted as well when (A) is under scrutiny. In fact, there is a case in which pluralism collides with unification. The concept that captures this situation is the "dis-unified theory". Literally it applies to (A) but (B) is indirectly under scrutiny, too. What does it mean for a theory to be dis-unified? Unfortunately, there are few definitions of fragmentation available. We can think of it as a reverse unification, usually called "fragmentation" of the domain of phenomena to which *T* applies. Intuitively, within *T* we discover a domain that is independent and has its own dynamics, distinct from *T*'s dynamics. This process is suggested by the pluralists, but I think it is better explained by one of the defenders of unification. N. Maxwell defends a thesis called the "comprehensibility of the universe" according to which: (1) the universe has two "aspects": *U*, present everywhere, throughout all phenomena and an aspect *V* that varies from place to place as determined by *V* and: (2) *U* is in principle knowable to us (Maxwell 1998 76-77). As presented by Maxwell, *U* is nothing else than the "unity of the world" discussed in the previous chapter. The universe is physically comprehensible when we have in principle a theory *T* of *U* and a dynamics of U. *T*, the theory of U needs to be unitary, not composed of parts (Maxwell 1998 77).[34] In characterizing *T*, Maxwell defines its degree of disunity *N* when the domain of phenomena which *T* applies to is "fragmented":

> for some (but not all) cases, the unity/disunity distinction can be indicated as follows. Let the candidate theory of everything, *T*, whose degree of unity is being assessed, predict possible phenomena *R.* If *T* is disunified to degree *N* then there are *N* distinct subordinate regions $R_1 \ldots R_N$ in the space of all possible phenomena *R,* different component theories, $T_1 \ldots T_N$, applying in each $R_1 \ldots . R_N$. For unity we require that $N = 1$. Different kinds of disunity arise depending on how the subordinate regions $\mathrm{R}_1 \ldots \mathrm{R}_\mathrm{N}$ are distinguished (Maxwell 1998 90).

[34] I do not plan to discuss Maxwell's more general project, but I intend to use his list of ways of dis-unifying a scientific theory for my present purposes.

Maxwell indicates no less than eight possibilities of fragmenting *R*, and consequently to disunify *T* (Maxwell 1998 91):

(1) By stating that $R_1 \ldots R_N$ are simply different regions of space and/or time;

(2) By stating that the range of a given dynamical law depends on the regions $R_1 \ldots R_N$ and on some other distributions of physical variables;

(3) Specifying that T applies to specific object(s) *O* when it/they is/are in a region R_i and not otherwise;

(4) Postulating a specific type of force for each R_i and maybe a universal force that applies to all R_i.

(5) Postulating different kinds of entity interacting by means of the same force; For each R_i there will be one type of particle to which a special type of theory T_i apply.

(6) Explaining *N* distinct entities as a result of a symmetry: “if the symmetry group is not a direct product of subgroups, we can declare that *T* is fully unified”; otherwise it lacks unity;[35]

(7) By spontaneous symmetry breaking, a pre-existing unity is broken and the current disunity, even apparent, cannot be explained by T;

(8) By postulating spacetime on one hand and particle and fields on the other hand. A theory that preserves the “spacetime-matter” duality is not unified.

According to Maxwell, by moving in this list from (1) to (5) the dis-unity is “less and less” severe such that in fact (8) is the least severe, but the most demanding to fulfill. Just by taking a quick look, I think that some fragmentations are *ad-hoc* enough to render

[35] See a similar discussion of T. Maudlin’s definition of unification in Section 12.6.

the resulting dis-unified theory as non-scientific: for example, almost any theory in physics are unified in denying explicitly (1), (2) or even (3). A problem with Maxwell's list is that theories in special science do not engage in strong claims such as (1)- (8) although they are dis-unified by other means. Even in theoretical physics, (1)- (8) are not widely accepted, contra what Maxwell suggests (Maxwell 1998 98). Another possible problem with Maxwell's list is that combinations between (1)- (8) are perfectly possible and Maxwell is not clear whether combining some of these procedures strengthen or not the fragmentation. Last but not least, there are several other aspects of a dis-unified science—not captured by Maxwell's list. My own proposal is to look at fragmentation and dis-unity at a smaller scale, on a case by case basis and not to the science as a whole. I endorse here the idea that different theories, even in theoretical physics, display a "local" independence in the sense that they evolve without commitments to claims such as (1)- (8).

There are more pragmatic reasons to endorse fragmentation. The dissenter would admit that each science is more or less accurate in a specific domain and that is it and all we can hope for. Focusing on the concrete particulars of all of these theories, advocates of this dis-unified picture say it is a dream to think it can all be told in one consistent theory. Instead of giving up unification, you can seek unifications that better conceal the new phenomena or unifications with as few as possible consequences which is indeed not a scientific enterprise anymore. This perspective of a "behemoth" science terrified several philosophers of science (most notably, I. Hacking).

2.2. Scientists against unified theories and unification

Unification is too cheap. Scientists' line of attack against unification echoes the worry that it is an artifact. Similarly to the philosophers who balk at unification, some physicists suspect that unification attempts are feigned mathematical hypotheses that force different theories under the same formalism without having any empirical support. Some arguments against fundamentalism and reductionism can be used against unification because, as we will see, reductive unification is one type of unification.[36] For fundamentalist, there are levels of interactions and some interactions are fundamental, while other are derived from those. The ontological fundamentalist is an eliminativist who wants to reduce objects at higher level of reality to objects at a fundamental level. In the last century the dominant attitude in physics was a species of eliminative fundamentalism. I take fundamentalism as a species of both ontological and epistemic reductionism. The fundamentalist hates dualities and he is always a fundamentalist about something. The epistemic fundamentalist draws the arrows of explanations (roughly, think of explanations as arrows from *explanans* to *explanandum*) as originating in one area of science, here fundamental physics. She thinks that there is a convergence of all the "arrows of explanations" toward one model, i.e. the model of elementary particle physics. The level of elementary particles is the fundamental one and the upper levels of organizations can be eliminated from ontology or from the epistemology of that theory. There is a fundamental science and all others are derivatives of it.[37] On one hand, the elementary particles *are* the fundamental blocks of the real world. On the other hand, relatively few laws provide all the information needed to

---

[36] See Chapter 5 on Morrison.

[37] I suppose St. Weinberg's strong fundamentalism displays a double fundamentalism: ontological and epistemological (Weinberg 1992).

represent the whole world and this information comes from fundamental physics. Epistemic fundamentalism by elimination of the non-fundamental level is clearly untenable when it comes to explanation or prediction.[38]

In attacking fundamentalism, some point out that there are theories which describe accurately enough some domains of physical phenomena, but there is nothing like a fundamental theory. One sometimes hears this too with respect to quantum field theory; namely, the idea that they might be effective field theories all the way down and that there is no final, fundamental theory (Castellani 2002; Cao and Schweber 1993). The theory itself is parceled and fragmented, without having the unity of laws or description one would like to see. Then the theory of everything is simply a conjunction of these domains without any coupling factor among them. For some, the artifice of a "fundamental theory" is too cheap in the sense that it is easily achieved and trivialized. Remember the so-called "notorious footnote 33" of (Hempel and Oppenheim 1948, 135-175) in which they had foresaw a major problem of the standard **D-N** model: the conjunction of the Kepler's law (k) and Boyle's law (b), *k&b*. Both *k* and *b* can be deduced in the sense of the **D-N** model from *k&b* (Salmon 1990, 3-24). Although it is formally correct, this is a trivialization of both explanation and deduction. This line of attack affects unification: the most trivial unifications are mere conjunctions of two theories: it is easy to see why a conjunction of Kepler's law and Boyle's law is spurious (Kitcher 1981; Maudlin 1996 131).

Feynman targeted unification when he used an example to mock some hyped-up attempts to unify all physical theories. Let us say that all laws have the form $A_i=0$ (for

[38] This is already a known argument by Putnam (Putnam 1979, 457).

example, $(F - ma)^2 = A_1; \left(F - G\frac{m_1 m_2}{r^2} \right)^2 = A_2 \ldots$, etc. and "the theory of everything" is: $\sum_i A_i = 0$ reminds against hyperbolized attempts to unification (quoted in Maudlin 1996; Feynman, Leighton, and Sands 1989 25-10-11). In order to have all the laws satisfied the "theory of everything" postulates that the sum of all As is zero only when all the laws are satisfied. In both of these mock cases, the "unificatory" theory makes no contribution (explanatory, confirmatory, interpretative on free parameters) in addition to the previous theories. A derivation of a law from the conjunction is a pointless "self-explanation" or "self-confirmation".

It seems that unification *tout court* is indeed a very boring and trivial idea. With no future constraints, one can see that unification by conjunction is in fact the only one with a general enough definition. The main question here is: what is then *the* real unification, if any? How do we sift the chaff from the wheat? I agree with some authors such as Teller and Morrison that for non-trivial cases, a definition is difficult to find.

Unification is too speculative and it is not corroborated with data. Other scientists, instead of mocking unification, think that it is too close to metaphysics and it causes more damage than one can imagine. During the last three decades an increasing number of physicists (R. Feynman, L. Smolin) and science writers (D. Lindley, J. Horgan) have balked at the state of theoretical physics and at scientific progress of theoretical disciplines (Smolin 2006; Woit 2006; Horgan 1996; David 1993). One common line of attack is that science is increasingly plagued by speculative theories, when it is linked to the "dream for unification". Fundamental physics seems "fundamentalist" in its quest for a final theory of everything. The Standard Model of particle physics (esp. the Dynamic Electroweak Symmetry Breaking), the Grand Unified Theory, Quantum Gravity, some cosmological models, and, above all, String Theory, claim unification as their main virtue. In recent criticisms of String Theory, for instance, one hears—among more particular worries—a complaint that physics need not be unified the way string theorists desire.[39] Witness the "superstring dream" as a full unification of one equation, one physical idea or one symmetry principle that explains everything (Smolin 2006 316; Woit 2006 261). As a preposterous speculation based on some metaphysical principle like "unity" or "beauty", theoretical physics is not science anymore and gets closer to mysticism and art based on the quest for beauty, symmetry and unity instead of truth. And the very idea of unification is the main culprit, when it is traded for empirical corroboration. Some empiricists suggest that empirical success trumps the non-empirical virtues such as unification, simplicity, generalization, mathematical beauty, etc. Whether truth and empirical success are related is a serious dispute that I do not want to address here. Without an empirical confirmation, a theory is deemed unscientific. Because of its lack of empirical confirmation, String Theory is usually appraised as "not ready for certain kinds of foundational studies" or still marred with structural problems (Weingard 1988; more recently Dawid 2006). More drastically, for R. Penrose and L. Smolin, String Theory already ended up by not describing the world and it cannot survive as a whole even the most modest confirmation (Penrose 2005).

There are quick answers to these worries, all coming from both the history and from a standard approach to philosophy of science. First, remember that in the history of physics

[39] L. Smolin, R. Penrose, C. Isham, etc., all tried this path as an alternative to the String Theory which is yet another unification.

some true models were blatantly disproven by data. They were empirically and explanatory unsuccessful theories, albeit true. My preferred example here is Aristarchus' heliocentric model (c. 270 BC), which was repeatedly falsified both by the data and by other models (mainly, the Ptolemaic model).[40] So here is the checklist of Aristarchus' model: problematic justification (the theory was wrongly justified on the basis of the Pythagorean philosophy that placed the fire at the center of the universe and not the earth), incomplete explanations (only Seleucus of Seleucia, c. 150 BC, correctly explained the tides based on this model without empirical support), precarious or no empirical confirmations (Aristarchus predicted the parallax but it was not possible to measure it, although astronomers tried hard to measure the parallax of distant stars; he predicted that the sun' diameter is 20 times bigger than the Moon's and 19 times farther from the Earth than the Moon; all of these predictions were in fact false), false predictions (mainly because of clumsy instruments and unreliable data). One can almost question the Aristarchus theory as being not a scientific theory. But Aristarchus' model was approximately *true* when compared to the Ptolemaic model. Aristarchus' idea needed eighteen centuries to resurface, precisely because its confirmation has been impossible, given the technology available. Similarly, Newton's corpuscular theory of light was not supported by experiment and it was not explanatory at that time, albeit in some respects it was closer to true than the wave theory of light. Empirical confirmation is not always a condition to the truth of a scientific theory.

Second, it is enough to mention here empirical and explanatory successful theories which were proven to be false: the Ptolemaic model of the Solar system, the phlogiston

[40] Aristarchus' work is translated and commented in (Heath 1913)

theory, the ether theory etc. Albeit false, they played significant role in the development of subsequent theories: parts of them were imported in newer and better theories.

Third, there is also a pragmatic dimension of confirmation. Some theories are difficult to corroborate, given current technologies and measurement techniques. We cannot measure or observe everything. Objects of study in cosmology, String Theory, Quantum Gravity, etc. are well beyond the energy limits of the present particle colliders and accelerators. We should not expect to deal with a simplistic confirmation theory in the case of these theories as we confirm a theory that operates at normal scale, let us say meteorology. As I will show, the same is true for Kaluza-Klein theories. In respect of confirming or disproving it, the theory is most likely false if taken in its literal interpretation. Empirical confirmation comes and goes and especially in the case of theoretical physics it is theory-laden: our methods and our instruments are therefore less reliable. It is well known that empirical confirmation is not the only criterion for doing good science. If one takes only the argument based on the under-determination of data by theories, our criteria to decide between theories cannot be restricted to empirical corroborations. Unification is something to be added to other theoretical/non-empirical virtues. Insisting too much on empirical confirmation is a bit naïve and based on a defunct logical positivism. Again, the contemporary debate surrounding String Theory are illustrative in this respect. N. Cartwright and R. Frigg suggested that in criticizing String Theory, physicists tend to simplify things: “this emphasis on prediction and experimentation is reminiscent of the philosophy of science of Karl Popper and the positivists of the Vienna Circle”. And proponents of String Theory exaggerate other virtues of it: simplicity, elegance, explanatory power. For Cartwright and Frigg String Theory should be judged on a multidimensional scale of criteria of scientific progress and not only empirical success:

> Although string theory has progressed along the dimensions of unifying and explanatory power, this in itself is not sufficient to believe that it gives us a true picture of the world. Hence, as it stands, string theory is not yet progressive because it has made progress only along a few of the many dimensions that matter to a research program's success. (Frigg and Cartwright 2006).

A bare empiricist would deny that we need to add anything to the empirical success when we choose our best theories, but such a simplistic position is easy to dismiss.[41] Since the dismissal of Logical Empiricism, it became clearer that it is impossible to do science without assuming something independent of empirical considerations. Confirmation is a complex relation because theories can get confirmed or, on the contrary, be disconfirmed by unexpected evidence. The fact that String Theory cannot be confirmed or disproved is related to the way it is formulated and to the scale it refers to (at the Planck length distances, durations and in colossal energy) and not to its claim to unification.

It is not clear what we are suppose to do when the empirical confirmation or disconfirmation is either missing or it is very unlikely to occur in the foreseeable future. I see the thrust of relating unification to confirmation, but for Kaluza-Klein theory we do not have yet the right experiments to confirm of disconfirm directly such a theory. There are several reasons, other than empirical or observational, to believe that Kaluza-Klein theories are literally false and I discuss some of them in Part III. My dissertation can be read as an attempt to emphasize the importance of non-empirical virtues in the genesis and evo-

[41] I suppose that one of the best illustrations of such a bare empiricism is Popper's *The Logic of Scientific Discovery* (1934). Later on, Popper himself would clearly admit that in choosing our best theories we need to apply to a "new requirement of simplicity" (Popper 1962 241).

lution of scientific theories. But there is an important caveat to the debate about non-empirical virtues: they have to be added to the empirical success and not be the only virtue of a scientific theory.

A general question raised is how important empirically uncorroborated, or even false models are for the progress of science. One cannot underestimate the importance of speculation and imagination in the evolution and progress of all sciences. Some speculations (String Theory included) might be useful for the progress of other sciences or for mathematical studies (Smolin and Woit admit happily this), even if they are not literally true. A skeptical question the empiricists should ask is: what scientific theory is literally true? In the absence of such a perfect theory we need to deal with “almost true theories” or with “approximately true theories”. My dissertation does not deal with this very convoluted issue. I take the truth of scientific theory as a relative measure that can characterize pairs of theories only. What is the metric associated to the approximate truth is difficult to establish in general, but a balance between empirical adequacy, predictive and explanatory power as well as *ad-hocness* can be the sought for metric. Although one can ask an interesting question whether the unificatory theory T is truer or not than the conjunction of T_1 and T_2, my dissertation is centered on the connection between unification and non-empirical virtues of theories such as explanation, simplicity, generality etc.

Unification is too expensive. Another concern of scientists is that once one achieves unification a different price needs to be paid: unification creates too much structure which: (a) is inconsistent with some other respectable theories, (b) does not meet certain condition of cohesion or (c) have unexpected empirical consequences, difficult to accommodate. Similarly to the concerns raised by some philosophers that unification creates “surplus structure”, scientists worry that empirical data is created. I call this “the empirical surplus” that comes with all unifications. Some scientists think that the unificatory surplus needs to be suppressed in order to avoid (a) and (b) and then the theory has to work its way out of (c), too. For example, deploring the attention and funding that String Theory received, L. Smolin makes some general claims about unification: “unification has *always* consequences which imply the existence of new phenomena” (my emphasis) and in many cases, these new phenomena “are not quickly seen or they disagree already with the phenomena.” (Smolin 2006 196). The suggestion is that unification comes with unexpected empirical consequences and with its own novelties that can rum sometimes against what we know about the world. In the case of successful unification these novel aspects of unification can be explanatory or can provide predictions. For other unifications, and this will be the case here, the novel consequences explain some facts which were taken as brute by previous theories. Smolin’s suggestion does not have a strong support in the history of science: unification comes with abnormal consequences but sometimes its consequences are predictive and explanatory.

2.3. The ideal of “modest unificationism” and a rebuttal

What if scientists such as Einstein or Poincaré were wrong? What if, *pace* Glashow, under the tapestry there is a messy enmeshment of unintelligible facts? What is then unification good for in the absence of a metaphysical unity? What bearing has this account of explanation if the world is dis-unified? Does unification imposes a unity where none is really to be found?

Unificationists such as Kitcher asked similar questions and answered some of them at a recent Symposium organized at the meeting of the *American Philosophical Association* on the Dis-unity of Science. In the modest, new "unificationism", unification is an "ideal" and has a strong normative component. We hope that the mathematical idealization deviates only slightly from real values; we hope that the study of organisms in a laboratory is at least slightly similar to their behavior *in vivo* conditions:

> Modest unificationists believe that the world may be a disorderly place, that the understanding of its diverse phenomena may require us to employ concepts that cannot be neatly integrated. Nevertheless, they see the practice of the sciences that I have just outlined as reflecting a regulative ideal, the ideal of finding as much unity as we can by discovering perspectives from which we can fit a large number of apparently disparate empirical results into a small number of schemata (Kitcher 1999 339).

For a modest unificationist like Kitcher, although many of these hopes are not yet fulfilled and we do not have a unificatory theory of everything, the parochial and fragmented practice of science is itself temporary. Kitcher would remark that despite present failures we should not infer that in the future a better theory will not discover a more unified treatment. The modest unificationist is idealistic and optimistic altogether. She sees here and there instances of scientific unification and merrily acknowledges them, whereas the big picture of science is still a messy and cumbersome one. "The local knowledge of today is a spur to the unification of tomorrow" (Kitcher 1999 348). His whole example which transpires the optimism of the modest unificationist is worth quoting:

> I imagine an eighteenth-century *savante* arguing that the particularities of inheritance and the details of the geographical distribution of organisms resist any attempts to assimilate them into the unified framework of existing science, that we just can't expect unified theories of inheritance and biogeography. With the advantages of hindsight we can tell her to wait a century for Mendel and Darwin (Kitcher 1999 342).

For Kitcher, at the end of the day, the dappled quilt would become a tapestry, *à la* Glashow despite what Cartwright advocated during the discussion. Kitcher admitted that other sciences (especially economics) resist any attempts to systematization and unification. In short this is his claim, call it the "Modest unity of the world":

> We hope that the world is unified; we need to organize our science "as if" the world is unified. Unity of the world is a regulatory ideal, not necessary a true one. We do not know whether the world is unified, but we hope so (Kitcher 1999 342).

For Cartwright, this "modest" optimism is not realistic. Scientists impose small pockets of unity by proposing mathematical principles or equations that are available for very restricted domains: "We may dream that the exact sciences will someday cover everything, but, I shall argue, that is not likely to be a dream that is even 'in principle' achievable." (Cartwright 1999a 319). The scientific endeavor works well only in cases of contrived experimental situations. We need to build nomological machines where Nature does not provide one. For Cartwright, these machines barely exist in nature and they have to be built in laboratories. In some respects, Cartwright will deny the global character of Kitcher's approach, albeit she might accept it locally. As an illustration, Cartwright lauded an economic model of A. Sen compared to that of Anand-Kanbur because the former is limited in scope whereas the second tries to infer too much and to expand beyond its own limits. As the history of science shows us, scientists who try to solve all problems by one stroke end up not doing science anymore.

For the dissenters, the claims of unity are in their majority boisterous, if not altogether dangerous for the practice of science. With some caveats such criticisms apply to unification, as well. I mentioned the pseudo-scientific explanations or unifications that

Marxism or Freudian psychoanalysis offered at the beginning of the previous century, those which had bothered Popper. By confronting these disastrous attempts to unity, a social scientist would naturally avoid unification and systematization as a global program in the special sciences. But this does not mean unification is not in principle possible. The modest unificationist sees unification as a good, regulative ideal, but does not intend to impose it in any case. Its instantiations are a question of contingency and cries for a deeper and more profound analysis on a case by case basis. Such a deflationary approach, mainly due to Margaret Morrison, is the topic of Chapter 5.

Chapter 3. Unification and explanation: M. Friedman and P. Kitcher

The core of my case study analysis is related to the connection between explanation and unification as developed by M. Friedman and P. Kitcher. I tackle this hypothesis in the earlier papers of Friedman (1974, 1981) and in Kitcher's subsequent improvements (1976, 1981). Later I enrich this discussion by coming back to causation in the specific context of explanation as discussed by Kitcher (1989,1990), W. Salmon (1989) and J. Woodward (2003).

There is a plethora of accounts of explanations: some of them are already dead, as it were, some were under scrutiny for decades.[42] Two accounts of explanation are still preeminent in the literature and relevant to the discussion on unification: the causal account of explanation and the unificatory account of explanation; by far, the former is the most popular.[43]

3.1. The D-N account of explanation

Intuitively, explanation is an operation by which something less familiar or not well understood is explained by means of something more familiar and better understood. *Explanandum* in Latin means the thing to be explained; *explicare* means literally "make level, flatten, or disclosing a thing", usually used in Ancient Rome in the sense of unfolding military tents during an encampment. The thing explained is made familiar by the way of

[42] I use here the term account to designate *philosophical* theories or models of explanation. I reserve theories and models as theoretical terms in direct reference to science.
[43] See a very recent account by M. Strevens who endorses a variant of the causal account of explanation based on difference-making, called the *kairetic* model (but alsoStrevens 2004; Strevens 2008).

another thing which *is* familiar, called *explanans* ("the thing that explains").[44] Darwin explained his concept of natural selection which was difficult to visualize and apprehend *in vivo* by the means of several examples of artificial selection performed by animal breeders, which anybody could easily understand. This is a relatively weak account of explanation because it carries us too close to subjectivity and relativism (Friedman 1974 9-11; Hempel 1965 430-3). Instead of making things familiar, explanations work perhaps as justifications to one's beliefs. Explanations are more or less arguments and provide reasons to believe *p*. The deductive character of explanations is obvious. The answers to question such as: "Why p?" take the form of an argument whose target is to persuade the person who asked the question. But this is again too weak. Persuasion would not count as a true explanation when the premises or the reasons are not justified themselves. Second, there are clearly beliefs for which "Why p?" cannot be pushed further on pain of a regress. We live in a world with brute facts and it is clear that some brute facts need to stay brute, *i.e.* unexplained. It is an interesting and difficult question to ask how we differentiate in principle brute facts from explainable facts.

Hempel realized that the "reasoning" model of explanation is deficient and added several important features to the reasoning account (Hempel 1965 364-376). Enter the deductive-nomological (**D-N**) account of causation, maybe once the most popular account in the philosophy of science—but not anymore. The **D-N** model is a deduction of an ex-

---

[44] The philosophical account of explanation can be traced back to—at least—Aristotle. Aristotle explained in the first book of the *Parts of the Animals* (*PA* 639b12-13) why causal explanations in terms of final causes are superior to any other types of explanations. This is why Aristotle's account of causal explanation is not taken as the being centered on teleology and not on causation *per se*. For a comprehensive discussion of teleology in Aristotle, see (Johnson 2005, 339).

planation from a set of premises such that at least one of these is a law of nature (Hempel 1965 335).

There is a natural trend to associate explanation to the search of a law of nature.[45] For Hempel, laws of nature are part of the scientific explanation. In his **D-N** model of explanation, an individual event *E* appears in the conclusion of a deductive argument. The premises are called the *explanans* and they are: 1) a set of lawlike statements L_1, L_2 … L_n and 2) a set of initial conditions C_1.. C_n. Event *E* is explained only if it is subsumed under a natural law.

$$\begin{array}{l} \text{Laws } L_1, L_2 \ldots L_n \text{ are established.} \\ \underline{\text{Facts/events } C_1, C_2 \ldots C_n \text{ occur}} \\ \therefore E \end{array} \tag{4}$$

The laws connect the *explanandum E* with the particular conditions C_1, … C_n and this confers the status of explanation.

Many explanations would simply fail to be **D-N** explanations because we usually do not directly employ laws of nature in explaining things. It is easy to debar such criticism by saying that the **D-N** account is in fact about scientific explanations and other explanations are “sketches of scientific explanations” (Hempel 1965 423). For example, the Doppler effect explains why the change in pitch of a train whistle when it passes in a railway station (from low to high and then back to low). We can use the difference in pitch to infer the speed of the train.

There are multiple problems that haunted the **D-N** model. First, witness the reliance of the **D-N** model on laws of nature. It has been argued that Hempel’s model is satisfied in

[45] The idea is present in Aristotle, too. For a reference from the heyday of this view, see (Suppe 1977).

neither natural nor social sciences. It needs universal laws which admit no exceptions. Hempel made a distinction between *laws of coexistence* e.g. Ohm's law or laws of ideal gases and *laws of succession*: Galileo's law, Newton's second law, etc (Hempel 1965 300). In laws of succession, time manifestly appears in differential equations: **PDE** or **ODE**. For Hempel, only laws of successions are deemed to be causal and laws of coexistence are not. This difference did not withstand multiple criticisms over time. But the **D-N** model and other nomothetic models of explanations depended on a distinction between laws and "accidental" generalizations. Even if one accepts that there are laws, how do we differentiate them from accidents that are true? This old Humean problem was one of the reasons to question the **D-N** model. In special sciences, although laws do not play a major role, still there is scientific explanation.

The other main problem with the **D-N** model is that it does not accommodate the theory of causation well. Explanations in particular sciences do not fit the **D-N** model well, as in many cases, even in biology, laws L_1... L_n are not present. Even in theoretical physics there are few cases in which we operate with L's. In many cases an idealized model is used instead of the law itself. On this account, we cannot make the difference between *explanandum* and *explanans* because sometimes the laws of nature are blind to such distinctions.

The literature mentions the well-known problem of symmetry. Bromberger (1967) pointed out that the **D-N** model can be used to explain both the length of the shadow of a flagpole at a certain moment of a sunny day and the height of the flagpole as a consequence, i.e. an *explanandum* of the shadow and the optics of the sunlight. Succinctly put, the **D-N** account leaves the class of the acceptable explanation of a given phenomenon underdetermined (Lipton 2004 28). Or, the **D-N** account lets us *explain* the speed of the train as a consequence in the variation of the pitch. These are of course not *bona fide* explanations. It seemed that something needed to be drastically reformulated in the **D-N** account. One proposal was to include causation in the account of explanation, mainly due to W. Salmon and D. Lewis (Salmon 1984; Lewis 1986).[46]

Hempel suggested that "often" the explanation of a particular event is conceived as specifying the cause. Causal explanation is a special type of deductive nomological explanation, although not all **D-N** explanations are causal (Hempel 1965 300). It also depends on the connection causation has with temporality, which is a major topic discussed in philosophy of physics. The **D-N** account was outlived by accounts that start from and rely on facts and relations among facts, most notably on causation. Salmon dubbed causal explanations "bottom up", while the "top-down" accounts of explanation are the **D-N** model and the "unificationist account" (both being epistemological). Salmon suggested that both top-down accounts are essentially deductive; indubitably, the great rifts in the **D-N** model of explanation have affected the accounts of unification.

3.2. Explanation and counting phenomena: Friedman (1974)

In the previous chapters I suggest that an analysis of unification *per se* is incomplete and that unification should be related to explanation, prediction, simplicity, *ad-hocness*, etc. The relation between unification and explanation has had the richest philosophical consequences. Unification exhibits relationship between phenomena or facts which were thought to be unrelated. Is this similar to the explanatory relation between

[46] The idea that in order to explain X we need to provide the causal history of that X does not belong properly speaking to D. Lewis. Aristotle and moreover Kant tackled in various ways the causal account of explanation in which understanding is related to the knowledge of causes. Lewis is important because he puts the causation in the context of regularities and counterfactuals which is definitively one of the most powerful accounts of causation (Lewis 1986), albeit problematic in itself.

explanandum and *explanans*? In fact, explanation was not used in order to elucidate unification, but the other way around: according to some philosophers, explanation is unification. So the answer to the question whether unification is related to explanation is couched in terms of necessary and sufficient conditions for explanation.

The first attempt to answer this question in the positive is attributed to Friedman who provided a theory of explanation based on unification. I call it *explanation qua unification*.

Before Friedman, various authors had suggested that explanation and unification are related. In 1918 Schlick suggested that by reduction, the number of things steadily diminish and consequently the number of explanatory principles used to attain the same level of knowledge is smaller. The ultimate task of knowing is to ascertain the fewest irreducible principles, i.e. those which are not susceptible of further explanation (Schlick 1918). W. Kneale suggested the same connection between explanation and the reduction of elementary facts:

> when we explain a given proposition we show that it follows logically from some other proposition or propositions. But this can scarcely be a complete account of the matter […] An explanation must in some sense simplify what we have to accept. Now the explanation of laws by showing that they follow from other laws is a simplification of what we have to accept because it reduces the number of untransparent necessitations we need to assume. […] What we can achieve […] is a reduction of the number of independent laws we need to assume for a complete description of nature (Kneale 1952 91-92).

Hempel took as a virtue of a theory its ability to explain an empirical law as an aspect of more comprehensive underlying regularities. For Hempel, such a theory provides a systematic unified account of many different empirical laws (Hempel 1965 444).

Friedman claims that all previous models of explanations, including Hempel's **D-N** model (or the "explanation as deduction" model), had failed to meet three criteria:[47]

- a theory of explanation should be sufficiently general: it should cover all scientific theories that we consider as explanatory.
- it should be objective: it should not allow a dependence on historical contingencies and fashions.
- it should connect explanation and scientific understanding: "We can find out what scientific understanding consists in only by finding out what scientific explanation is and vice versa" (Friedman 1974 6).

As none of the former theories of explanation had satisfied these conditions, Friedman's explanation *qua* unification was designed specifically to meet them:

> I claim that this is the crucial property of scientific theories we are looking for; this is the essence of scientific explanation—science increases our understanding of the world *by* reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things equal, more comprehensible than one with more (Friedman 1974 15).

In short, Friedman defined explanation in terms of unification. One should immediately notice that Friedman's account hinges on the idea of "independent phenomena" and on "reducing" their number. In order to work this out, he presupposes a deductively closed set of law-like sentences *K* and their consequences: "[…] we can represent what I have been calling phenomena i.e., general uniformities or patterns of behavior—by law-like sentences; and that instead of speaking of the total number of independent phenomena we can speak of the total number of (logically) independent law-like sentences" (Friedman

[47] Other popular models of explanation in the 1970s were explanation "as reduction to familiar terms" of M. Scriven, W. Dray and S. Toulmin's "explanation as cultural pattern".

1974 15). But here comes a difficulty formerly discussed by Hempel and Oppenheim: how do we count the conjunction of two laws S=K_1&K_2? Is S a "third" independent law? Intuitively, S reduces the number of independent sentences in K, but the account cries for a definition of independence of two sentences:

> ***[4]*** *P_1 is independent of P_2 iff the grounds for accepting P_1 are not sufficient grounds to accept P_2.*

If $S \vdash Q$, then S is not acceptable independently of Q and if S is acceptable independently of P and $Q \vdash S$, then S is acceptable independently of Q (Friedman 1974 15). Aside from this, a concept of partition associated to a sentence S is necessary. A partition $\Gamma = con_K(S)$ is a set of sentences logically equivalent with S *and* which can be accepted independently of it. There are K-atomic sentences in the sense that they have no partition. A clearer example of partition is provided by W. Salmon: Newton's law of universal gravitation is partitioned as follows:

> Between all pairs of masses in which both members are of astronomical dimensions there is a force such that gravitation, (2) Between all pairs of masses in which one member is of astronomical dimensions and one is smaller there is a force such that gravitation, (3) Between all pairs of masses in which both are of less than astronomical dimensions there is a force such that gravitation. (Salmon 1990 6).

In order to give a formal meaning of "reducing the number of independent sentences" Friedman employs the idea of the K-cardinality ($card_K$) of a partition:

> ***Def 2*** A sentence S reduces the set Δ iff $card_K(S \cup \Delta) < card_K(\Delta)$

By this, Friedman solves the conjunction paradox. Indeed, the conjunction of some law $S = L_1 \,\&\, L_2 \,\&\ldots$ does not reduce the set of $\{L_1, L_2 \ldots\}$ in the sense of **Def 2**.

The final stage of Friedman's construction is the definition of explanation. The first requirement is that the explained sentence is to be in the partition of the *explanans*. A sentence S_1 explains another sentence S_2 if S_2 is an independent consequence of it, i.e. if it belongs to $con_K(S_1)$. But this is not sufficient. The second part of the definition of explanation demands of S_1 to *reduce* $\mathrm{con}_K(\mathrm{S}_2)$.

Def 3 S_1 explains S_2 iff:

1) $S_2 \in con_K(S_1)$

2) *S1* reduces $con_K(S_1)$.

This definition can be weakened to:

Def 4 S_1 explains S_2 iff: there is a partition Γ of S_1 and a sentence in it $S_i \in \Gamma$ such that $S_2 \in con_K(S_i)$ and S_i reduces $con_K(S_i)$

Friedman claims that **Def 3** and **Def 4** are no more vulnerable to the conjunction objection.

Here are some of the differences between Friedman's book (Friedman 1983) and the two papers discussed in this section (Friedman 1974; and Friedman 1981). First, the chapter in the book directly relates unification to confirmation and then to realism, whereas the first paper (Friedman 1974) is centered on how unification is related to explanation. So the stake is higher in the book chapter. What makes unification desirable in the book is its connection to confirmation rather than the relation to understanding. In some sense, the 1983 account is opposed to explanation, especially to the principle of the "inference to the best explanation" (**IBE**), because explanation does not provide a guide to the interpretation of theoretical structures. Friedman tried to show that we cannot limit the physical world to the observational world and that we need to associate parts of the theoretical structure to the world—in a literal sense. Some theoretical entities are paradigmatically problematic: for example, the absolute Newtonian spacetime which does not literally exist. Some other theoretical entities exhibit unificatory power without being problematic. Again, how do we separate the wheat from the chaff? Friedman suggests that the inference to the best explanation does not work in this case and it should be replaced with the unificatory power of the theory. The main question is whether we interpret theoretical structures literally or instrumentally. Given an observational structure $\mathcal{B}$ one can always find a theoretical structure $\mathcal{A}_1$ such that $\mathcal{A}_1$ models $\mathcal{B}$. According to Friedman, we can also create another structure $\mathcal{A}_2$ that models better $\mathcal{B}$, and so on. The "inference to the best explanation" will never stop and progressively it becomes trivial. Even when $\mathcal{A}_1$ is the only structure available, "the best" as in "the only one available", may nevertheless be an unsatisfactory explanation (Friedman 1983 259)**.** The whole debate surrounding the inference to the best explanation (**IBE**) misses the point here: even for one pair of structures $\mathcal{A}$ and $\mathcal{B}$, there will be always two interpretations available: a literal and a representational one. **IBE** cannot provide an answer to the question why theoretical structure should ever be taken literally.

Friedman claims that he managed to replace the "best available explanation" with the notion of unificatory power: "a theoretical structure should be taken literally when, and only when it has sufficient unifying power." (Friedman 1983 259)**.** Hence the strong connection between realism and unification. But in order to decide about the unifying power of a theory, confirmation is essential.

The second difference is that whereas Friedman's articles (Friedman 1974; Friedman 1981) had been completely couched in the syntactic view of scientific theories for which theories are collections of sentences, in the book, Friedman approached unification differently (Friedman 1983). Here theories are not mere collections of sentences, but the internal structure of theories comprises terms and relations and the relation is between the model and the sub-model. Margaret Morrison's criticism relies on this difference (Morrison 2000 37). Third, the way Friedman deals with the conjunction of two theories is different in both approaches: the discussion in the book applies to both sorts of unifications, whereas the conjunction of two theories does not qualify as unification in the first paper (Kukla 1995 236).

3.3. Patterns of argumentation: Kitcher (1976, 1981)

In the early approach to unification, P. Kitcher put a new emphasis on the link between explanation and unification by providing counterexamples to Friedman's account (Kitcher 1981, 507-531; Kitcher 1976). There are *bona fide* conjunctions which have explanatory purchase. For example, the adiabatic expansion law is explained by a conjunction of the Boyle law and the first law of thermodynamics. In dealing with a complex system we need conjunctions of laws from different areas (the explanation of how an organism works needs conjunctions of laws of physics, chemistry, biology, etc.). For Kitcher, **Def 4** does not accommodate non-trivial conjunctions. His early solution is to employ *kinds* of laws, not simply sets of laws as Friedman did (Kitcher 1976 212). But the more developed model is deployed in a later article (1981) where Kitcher relates explanations back to arguments as in the DN model, more precisely to *patterns* of arguments:

"Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same patterns of derivations again and again and, in demonstrating this, it teaches us how to reduce the number of types of facts we have to accept as ultimate (or brute)" (Kitcher 1981 432; revisited in Kitcher 1989 esp. §4)

Explanations are methods of answering "why-questions" such as "why is it the case that *S*?" where *S* is the conclusion of an argument. *Principia*, for example, uses *kinds* of argument applied to the force of gravitation. Newton intended to extrapolate them *formally* to other forces than gravity (by the so-called "corpuscular dynamics"). This is true also for Darwin's *explanation sketches* of evolution. These two historical examples led Kitcher to the conclusion that "the notion of an argument pattern is central to that of unification" (Kitcher 1981 510). The way of dealing with explanations *qua* arguments is to define an argument pattern as "a triple consisting of a schematic argument, a set of sets of filling instructions containing one set of filling instructions for each term of the schematic argument, and a classification for the schematic argument." (Kitcher 1981 515). A schematic argument is a sequence of schematic sentences. A schematic sentence is a logical expression obtained by replacing some nonlogical expressions (but not all) in a sentence with letter symbols. Arguments patterns are not expressible in simple sentences, but in schematic sentences in which the nonlogical terms are replaced by (dummy) letters. The filling instructions for a schematic sentence are specifications for replacing the letter symbols of the schematic sentence. A classification is a description of the inferential characteristics of the schematic argument: "its function is to tell us which terms in the sequence are to be regarded as premises, which are to be inferred from which, what rules of inference are to be used, and so forth" (Kitcher 1981 516). A general argument pattern is instantiated by a set of sentences if some conditions are satisfied: (i) The sentence sequence and the schematic argument of the general argument pattern have the same number of terms; (ii) All sentences in the sequence are obtained from the schematic sentence in accordance with the appropriate set of filling instructions. (iii) a chain of reasoning which assigns to each sentence the status accorded to the corresponding schematic sentence by the classification can be build up.[48]

Kitcher employs a less formal notion of similarity between patterns of arguments. The punch line is that scientists are using "similar" arguments in the derivation of many accepted laws in which some non-logical terms are kept and some are replaced. The pattern imposes non-formal conditions on arguments which instantiates it. (1) On the one hand, there are restrictions of the instructions for replacing letter symbols because of the non-logical terms present in the pattern. (2) On the other hand, the classification imposes conditions on the logical structure. If both conditions are relaxed, then the notion of pattern degenerates. If both conditions are enforced, then the pattern instantiates only itself and unification becomes trivial. Kitcher calls "stringent pattern of argument" those in which the non-logical terms impose strict requirements on their instantiations. In this sense a stringent argument is more constrictive than a formal one without non-logical terms. If we take a domain of knowledge K as a "set of sentences endorsed by the scientific community" the systematization of K is a set of arguments that are acceptable to it (Kitcher 1989 431).[49]

---

[48]An example is of an argument instantiating an explanation is: "the force on α is β; the acceleration of α is γ; force = mass · acceleration; (mass of $\alpha \cdot \gamma$)=β ; $\delta = \theta$" where α is to be replaced by a body, β is a function of coordinates and time, γ is the acceleration as a function of coordinates and time, δ is an expression of a coordinate and θ is a function of time (Kitcher 1981 517).

[49] I.e. all their premises and conclusions are in K.

E(K) is a set of arguments that is the best systematization of the domain K. This is the "explanatory store" of K.

If Σ is a set of arguments, then a generating set for it is the set of patterns Π instantiated by Σ such that each member of it is instantiated in Π. Giving a set Σ_i of arguments acceptable relative to *K*, its generating sets are Π_{ij} (obviously, there are various generating sets for each set Σ_i). From these generating set one can extract a basis B_i associated to Σ_i as the best from Π_{ij} in respect of unifying power. If B_m is the basis with the greatest unifying power, then the explanatory store $E(K)=B_m$.

If $C(\Sigma)$ is the conclusion of Σ, (C is a set of conclusions of some arguments in Σ), then at a first sight, the unifying power of a basis B_i with respect to K "varies directly with the size of $C(\Sigma_i)$, [varies directly] with the stringency of the patterns which belong to B_i and [varies] inversely with the number of members of B_i". A partial conclusion would be that "unifying power is achieved by generating a large number of accepted sentences as conclusions of acceptable arguments which instantiate a few, stringent patterns" (Kitcher 1989 520) But this is unsatisfactory for Kitcher. And this is another difference between Kitcher and Friedman: instead of merely "counting" the different patterns in a basis of arguments, his proposal is to look for *similarity* among them.

A first virtue of Kitcher's account of explanation qua unification is that it can solve the "irrelevance problem", the "asymmetry problem" and the accidental generalization that were familiar difficulties to the covering law model. He claims that the account can debar spurious unifications cases. How can we rule out explanations based on the following argument:

$$\frac{\alpha \,\&\, \beta}{\therefore \alpha} \tag{5}$$

Here the criterion of stringency is useful. The argument succeeds in generating many beliefs by quite a few patterns, but it fails the criterion of stringency, more exactly the constraint (1) above, i.e. it is too lax in allowing any vocabulary to appear in the place of α. If one adds some pseudo-restriction on the replacement of α, then the allegation of spurious unification is still legitimate. To this end, Kitcher employs a new constraint to rule out spurious unification: “if the filling instructions associated with a pattern P could be replaced by different filling instructions, allowing for the substitution of a class of expressions of the same syntactic category, to yield a pattern P’, and if P’ would allow the derivation of any sentence, then the unification achieved by P is spurious” (Kitcher 1989 527-8). In Kitcher’s rendering of the new condition, spurious unifications as given by the above argument and another class of arguments as:

God wants it to be the case that α.
<u>What God wants to be the case is the case</u>. (6)
Therefore, α.

is also ruled out as spurious unifications. Nonetheless, there are ways to sneak in some spurious unification in Kitcher’s model and he acknowledges this. The main reason is that his definition of stringency is obviously incomplete. But some of them can be eliminated in this manner.

Chapter 4. Causal explanation versus unificatory explanation (Kitcher, Woodward)

As in the case of unification, there are some skeptics who question the strong connection received view that explanation is always couched in terms of causation. Does causation always provide explanation? Even if one knows the complete pattern of causes that produced an event, one does yet not have a scientific explanation. We prefer *pieces* of causal chains or better, those pieces of causal chains that are *relevant* in a given context (Railton 1981). Can unification constitute an alternative to the causal explanation?

4.1. Unificationism: explanation without causation (Kitcher, 1989)

In trying to address the relation between unification and causation, in a collection co-edited with W. Salmon, *Scientific Explanation (Minnesota Studies in the Philosophy of Science, vol. XIII)*, Kitcher suggested a more radical approach to unification in which explanation is unification in the sense that causal notions are derived from explanatory notions: "If F is causally relevant to P, then F is explanatory relevant to P" (Kitcher 1989 495). For Kitcher the causal asymmetry, present in examples such as the flagpole and its shadow, unveils a basic asymmetry that involves unification.[50] Explanation *qua* unification has been classified as an alternative to the standard model of explanation *qua* causation such that a possible conflict between explanations based on causation and those based on unification is suggested.

[50] Some philosophers think that Kitcher's original motivation to develop the unificationism account was his distrust of the idea of causation (Godfrey-Smith 2003 196).

According to M. Strevens and to S. Psillos, i.a., there are two types of accounts of explanation: the unificationist (Friedman and Kitcher) and the causal approach (adopted by the majority of the philosophers of science; maybe W. Salmon and J. Woodward best illustrate this commitment to causation in approaching explanation). I do not suggest that unification is absent in Salmon's standard account of causal explanation. On the contrary, in the last section of his 1984 book, Salmon confesses that:

> The ontic conception looks upon the world, to a large extent at least, as a black box whose workings we want to understand. Explanation involves laying bare the underlying mechanisms that connect the observable inputs to the observable outputs. We explain events by showing how they fit into the causal nexus. Since there seem to be a small number of fundamental causal mechanisms, and some extremely comprehensive laws that govern them, the ontic conception has as much right as the epistemic conception to take the unification of natural phenomena as a basic aspect of our comprehension of the world. *The unity lies in the pervasiveness of the underlying mechanisms* upon which we depend for explanation (Salmon 1984 276).

For Salmon unification is a contingent concomitant upon the discovery of the causal structure: sometimes unification comes in the same package with causation, sometimes it does not. For Salmon, it is perfectly possible that a factor F that is causally relevant to some phenomenon P does not appear/occur in any derivation occurring in the explanatory store that explains P. F could be present in the premises of no argument whatsoever that infers P. In general, for many others, unificationists are too superficial and they miss the point: "One could think that the very fact the unifying argument-patterns are those that preserve the intuitive asymmetry in the order of explanation points to a deeper characteristic they have; namely that the capture facts about the *causal order* of the world" (Psillos 2002 278). But as M. Strevens pointed out, the unificationist is able to handle easily cases where the causal structure seems to matter the most, by showing that:

> the omission or distortion of the details enables a far greater degree of unification than would otherwise be possible. The causalist is embarrassed, because giving the correct causal details is, on the causal approach, just what explanation is supposed to be" (Strevens 2004 156).

I take Strevens' dissatisfaction as a central tenet of other authors who criticize the Friedman-Kitcher approach. The division cuts deeper than the discussion surrounding unification and it does not involve unification in an essential way. A Humean could stand up for Kitcher's cause and claim that causal order does not precede the explanatory order. Indeed, as unificatory theories relate primarily to the latter, why should we impose causation upon unification? Kitcher claims that the causal order is understood either in terms of the concept of explanation or "in terms of concepts that are themselves sufficient for analyzing explanation" (Kitcher 1989 420). The opposite view, illustrated by Salmon, Woodward *i.a.*, simply cannot accept that explanation comes before causation. Kitcher himself admits that individuals do not go through the grim process of comparing competing deductive systematizations with respect to number of number and stringency of patterns and number of conclusions in order to determine which is the most unifying. Most people absorb the "lore" of their community which is in most cases a causal one (Kitcher 1989 436). One cannot infer from this that the lore is the end of story though.

Kitcher started off with a critique of van Fraassen's proposal that explanation are not arguments, but answers to "why" questions. I already showed how easy is to trivialize unification: likewise, explanation can be trivialized. Kitcher and Salmon showed that van Fraassen's pragmatics is not enough because it can be trivialized in the sense that *any* true sentence A will explain *any* sentence B in a context K={B,~B} (Kitcher 1989 415). One of the problems of the **D-N** model was that it could not solve the asymmetry problem. If the length of the shadow of a flagpole is explained by its height (and other relevant facts related to optics, sunlight, etc), how is that the height of the flagpole is not explained by the length of its shadow? Indeed, for mostly all imaginable, simple cases, the length of the shadow is not relevant to its height. Apparently, the **D-N** model misses here the relevance relation between the length and the shadow of a flagpole, which is the causal relation. *Pace* Kitcher, an explanation that is not able to render the genuine, causal relation is not really an explanation. The reason is that the notion of law of nature cannot be grounded in that of cause. This is music for the ears of a Humean, too, who would prefer to use explanation in an analysis of causation, not the other way around. In this respect, as I mentioned previously, it is perfectly reasonable to be pluralistic in respect of explanation and admit that there are explanations that are not causal.

In his attempt to address similar worries in the case of unification, Kitcher remarks that in mathematics or in formal syntax there are explanations of facts that do not have causes (Kitcher 1989 423). Geometrical explanations or more general mathematical explanations are not causal, but are they explanations at all? Philosophers who embrace somecausal explanation would suspect that there are explanations based on geometry or on mathematics in general are less fundamental and less important than causal explanation. Mathematical explanations are deductive, as unificatory explanations are. This is suggested by J. Woodward and M. Morrison, see Section 4.2 and Chapter 5). I will deal with this argument in Part III and it will be the thrust of the discussion in Part III. I am less concerned with explanations within mathematics, but with mathematical explanations in natural sciences. For some philosophers, mathematics is the source of all explanations: Shapiro claims that “a scientific ‘explanation’ of a physical event often amounts to no more than a mathematical description of it" (Shapiro 2000 34) Shapiro realizes that we need a clear account of how mathematics is related to scientific practice. The question that can be asked in a more general context is whether mathematics is indispensable or not to scientific theories.

One of the major difficulties of causal explanation is its epistemology; Kitcher refers here to Salmon (Salmon 1984 305; Kitcher 1989 460). How do we gain knowledge of the causal structure of the world? The causal approach faces the difficulty of formulating necessary and sufficient conditions for obtaining causal relations. But is causation observationally ascertainable? The popular idea exploited by Salmon is Reichenbach's: causal processes transmit information because they are markable. Kitcher disagreed and spent a great amount of time in exposing the major troubles of the marking theory of causation and of one of its alternative, the counterfactual theory of causation (Salmon 1984 142; Kitcher 1989 461-475). Notwithstanding all these troubles of causation, Kitcher is optimistic in respect of explanation and bolsters explanation as being able to ascertain progress in knowledge. But instead of being a deployment of causal connections, explanation is a systematization of beliefs via argument patterns. Subsequently, this serves as the basis for the introduction of causal concepts. "[…] the 'because' of causation is always derivative from the 'because' of explanation. In learning to talk about causes or counterfactuals we are absorbing earlier generations' views of the structure of nature, where those views arise from their attempts to achieve a unified account of the phenomena" (Kitcher 1989 477).

The bottom up strategy adopted in the causal explanation approach is premised on the idea that we can discern causal relations in specific phenomena and then construct from them theoretical explanations. For Kitcher, in the opposed strategy—the top down approach—theoretical explanations are used to account for causes in specific cases. To know a theory does not mean to answer the why questions it asks, but to "internalize its argument pattern". In an example taken from classical genetics, Kitcher showed how a single pattern of derivation can be used to derive a variety of conclusions (Kitcher 1989 sections 4.6.1-4). Kitcher claims that his unification account is able provide a genuine methodological guidance in deciding which theory is better, unlike the causal account of explanation.

A legitimate question is whether the unification account can surpass the difficulties of asymmetry and irrelevance that marred the **D-N** model of explanation.

Kitcher's last aim is to provide an account of scientific progress via unification. Scientists follow an *origin-and-development* i.e. a temporal story of the origin of the objects involved. Let K be a set of beliefs and E(K) its explanatory store. An argument S that runs from the premises about the length of the shadow of a flagpole to its height contains an extra pattern than E(K) and in that respect it fares worse than E(K) (Kitcher 1989 485). Given two languages L and L' and two sets of beliefs K and K', we want to know whether there is an explanatory gain from $\left\langle \mathrm{L}, \mathrm{K}, \ \mathrm{E}(\mathrm{K}) \right\rangle$ to $\left\langle \mathrm{L'}, \mathrm{K'}, \ \mathrm{E}(\mathrm{K'}) \right\rangle$. If K' employs the same stringent patterns to generate more consequences or, alternatively, more consequences from the same stringent patterns, then E(K') unifies K better than E(K) unifies K.

I see a major difficulty in Kitcher's account related to the way in which we count different phenomena. Is, for example, the trajectory of the cannonball different than the orbit of the Moon? Are these two phenomena explained by Newton's theory or they are in fact one? The answer seemed to be in the positive, but for a geometer, even a pre-Newtonian geometer both were conics. For an Aristotelian metaphysician, they are totally different because of some prior beliefs related to the impossibility of a mathematical description of celestial bodies. It is clear that once we shift to a different metaphysical framework, things start to look either very different or very similar. The same can be said about two objects with the same mass and the same mechanical features which will follow two similar or identical trajectories, let us say (a) a rock; and (b) a bird. Or let us say the descriptions of (c) an interaction between a photon with a wavelength of 510nm and a detector; and (d) between a photon with a wavelength of 511nm and the same detector? Or an experiment with (e) an elementary spin-1 particle or (f) a spin-2 particle. Are cases (a) and (b) similar or different enough in Kitcher's account of unification? What about cases (c)-(d) and (e)-(f)? Do these phenomena count as generating different conclusions of the same pattern of; let's say Newtonian mechanics, and **QM** respectively? Counting different phenomena as conclusion of the same stringent argument is a difficult business for Kitcher and it seems that it is relative to a system of background knowledge or background assumptions as well as to some set of subjective beliefs. Similar points can be found in (Morrison 2000; Woodward 2003 366; Salmon 1989). In comparing the explanatory store of two theories, a possible strategy is to count *types* of facts not individual facts. This strategy has its own difficulties, too: the type of phenomena is not in itself objective, but it depends on a scientific theory or on metaphysics.

Another difficulty that Kitcher faces here is to account for the comparison of patterns that are formulated in different languages. But instead of taking it as a problem, Kitcher holds that explanatory unification can serve as the basis of a defense for conceptual change. The proviso is that the change from K to K' and the change from L to L' are "defensible" in the sense that there are no strong arguments from the perspective of

$\langle \mathrm{L}, \mathrm{K}, \ \mathrm{E}(\mathrm{K}) \rangle$ against the shifts envisaged (Kitcher 1989 491). K needs to be neutral toward the changes to K'.

4.2. Woodward's three criticisms against unificationism

In defending the causal model of explanation, Woodward recently raised several criticisms against Kitcher's 1989 unificationism account of explanation (Woodward 2003 Ch. 8). Woodward's analysis echoes his own interventionist view of causation and I take as one of the most elaborated account of causal explanation.

(A) For Woodward, unification is a very complex type of scientific achievement. A general account of unification cannot capture all of them. There are so many different cases of unification, some stronger and some weaker. Some are strongly committed to explanation, some not at all. For Woodward three types of unification are significant:

(I) The "classificatory unification" is present when a vocabulary is able to better classify a wide range of phenomena; an example is the Linnaean taxonomy in biology. According to Kitcher, belonging to a specific category, e.g. mammals, gives information about other predicates of the individual (having backbones, heart, etc).

(II) unification can be achieved by a common mathematical formalism that applies to a huge variety of phenomena: Woodward's examples are the Lagrangian and Hamiltonian formalisms in classical mechanics. In fact, the treatment of physical processes in terms of **PDE** unifies almost all sciences.[51] This is the "mathematical unification" on which I will come back in the next chapter.

[51] I do not enter here in the debate whether the world at the fundamental level can be described by **PDE**. A vast part of our world can be described by **PDE**s.

(III) unification is achieved when phenomena "previously regarded as having quite different causes or explanations are shown to be the result of a common set of mechanisms or causal relationships" (Woodward 2003 362).[52] (III) is the "physical unification" in Woodward's account.

For Woodward, the enthusiasts for unification should not fool themselves that by adopting a common classificatory language or a common mathematical framework scientists unify a wide range of phenomena. It is difficult to deal with all these three types in a single stroke; Woodward opines that only the latter type has anything to do with explanation and the other two may not be related to scientific explanation at all. (I) and (II) have nothing to do with explanation and they are tricky enough in themselves. In respect of (I), it seems that there is virtually countless ways of classifying living things; there are also infinite ways of formalizing different phenomena.[53] Woodward wonders whether Kitcher's account is able to discriminate (I) and (II) from (III). How do we separate causal unification, i.e. the physical unification (III) from other sorts of non-causal unification, (I) and (II)?

Type (II) is more relevant for the present dissertation. It is true that natural language and mathematics facilitates the communication among scientists; helps them to share results or procedures, simulations, models etc.. Mathematics is in many cases "already there" when scientists discover a new law or a new phenomena. Lagrange equations are able to describe the patterns of behavior of many types of systems which do not involve any kind of common set of causal or explanatory factors (Woodward 2003 363-4). Lagrange equa-

[52] Woodward talks here about Newton's unification.

[53] Many of these unifications are simply trivial or irrelevant. For Woodward, Linnaeus' biological classification is not explanatory because there is no intervention by which we can change an organism into a mammal or a polar bear (Woodward 2003 363-4).

tions are used to derive predictions about both the behavior of a system of gravitating masses and an electrical circuit. It would appear that there is a straightforward sense in which this involves the use of the same general pattern of derivation over and over again and hence that this ought to count as an explanatory unification. […] However this "unification" does not seem to involve a common set of causal or explanatory factors (Woodward 2003 365). And as I already indicated, I concur with Woodward that explanation is part of the aim of science, despite what others would say—mainly the empiricists, L. Sklar or J. Norton who think that explanation in science is overstated. For me, explanation plays a specific role in theory-choice as well as in getting over the problems of underdetermination.

I have an answer to Woodward's first challenge. There are very few fortunate cases when the mathematical formalism matches the causal structure. In general, the problem is that a system of equations with a definite solution can be represented as an infinite number of regression equations. In other words, to a dynamics we can associate innumerable causal structures. What is needed is a further property called "modularity". In order to obtain it we need to decompose the system under scrutiny in parts and their mechanisms: "understanding the behavior of a complex system is a matter of representing it as segregated into parts and components, where the representation is modular in the sense that those components are represented as changeable independent of each other" (Woodward 2003 337). Modularity has to be added to the mathematical formalism in order to explain. For Newtonian mechanics, this cannot be done because gravitation is action-at-distance. In general, the physics of a classical system contain a great deal of irrelevant information and may well omit the relevant information if this causal information is not added.

(B) Woodward echoes a worry—expressed by several authors—that Kitcher's approach needs a winner-takes-it-all conception according to which generalizations or theories cannot be explanatory without being unificatory. The winner-takes-it-all conception would lead to comparisons among explanations based on their unificatory power such that at the extreme, in a given science, only the most unified theory is explanatory and everything else is non-explanatory. Generalizations in science have to be unificatory or else they do not explain at all. For this strategy, explanation and unification go together in the sense that "theories that are less unifying than the most unifying *known* theory are not explanatory" and that "only theories that are explanatory are the most unifying theories that will ever be discovered (or perhaps the most unifying theories that exist, whether or not they will ever be known)" (Woodward 2003 368). There are some undesirable consequences of this claim of the unificationist. She has to infer that a theory T_1 is explanatory at a moment of time in virtue of the unification it has achieved, but it becomes non-explanatory once another more unifying theory T_2 becomes known. This entails that we know the degree of unity of a theory and we are able to compare the explanatory stores of theories. In other words, we need to know what the explanatory stores of the two competing theories are and how unificatory they are in order to compare them. One problem is that in many cases we cannot in fact estimate the degree of unity of our current theories. The second is that examples of non-unificatory theories that nevertheless explain abound in science: Coulomb's and Ampere's theories are explanatory although they are less unificatory compared to Maxwell's. Woodward suspects that the practice of science does not consist of an evaluation of the justification patterns and their stringency, i.e. of their unificatory power, but of a constant search for causes. The process of comparing various degrees of unification is unconsciously carried along with the conscious process of acquiring new causal knowledge (Woodward 2003 369).

I want to give two short answers to Woodward. First, it looks to me difficult to see why we discover causes easier than we estimate the stringency of arguments. Is the theory of evolution about discovering causes more than discovering an argument pattern? Or did Maxwell look for "a causal agent" when he discovered electromagnetism or, on the contrary, was he looking of a stringent argument? Second, the unificationist needs a weaker condition: she needs to compare the degree of unity of theories instead of knowing the absolute "unity value" of a given theory. Here my strategy of separating unity of theory from its being unifying as the result of a unificatory process proves to help the unificationists: we can and we do compare the unificatory power and the explanatory power of a unifying theory T with its less unified predecessors. We do not want to compare any two scientific theories, but those theories which were unified with the theory that unifies them.

(C) There is a general worry that unificationism is "deductive chauvinistic".[54] For Kitcher, all explanations are deductive and purported non-deductive explanations can be construed as deductive arguments. What if some explanations are not deductive? Woodward's suspicion is that the unificationist cannot escape the asymmetry problem raised against the D-N model. Instead of reconstructing arguments, scientists look for answers to "w-questions". For interventionists, generalizations are related to range of invariance rather to their scope.[55] Unificationism is based on scope and not on invariance. Some generalizations can have a wide scope without being invariant and the other way around

[54] This is how Kitcher labeled his own unificationism (Kitcher 1989 sect. 5)
[55] See also (Woodward 2003 Ch 5 & 6)

(Woodward 2003 367). But for the interventionist, deductive arguments, mathematical manipulations, formalisms, etc. do not necessarily trace or represent causal relationships. Woodward's theory is centered on the idea that "the causal order, as reflected in facts about the outcomes of hypothetical experiments, is independent of and prior to our efforts to represent it in deductive schemes" (Woodward 2003 361)**.** Unification in itself does not select explanatory derivations from non-explanatory derivations.

4.3.Two replies to Woodward: diversity in explanation

It is important to clarify a general point about Woodward's account. In his book he wards off mathematical and descriptive explanations. According to the more liberal and pluralistic view, descriptions, classification and mathematical formalisms are all able to explain. For example, for Paul Churchland explanation is associated to any activity of recognition, classification, or description that we perform in everyday life (Churchland 1989, 321). On the contrary, Woodward does not intend to capture those types of explanations and he narrows his theory of explanation to causal explanations. It is important to make this point in the context of Morrison's analysis, who is less clear on her theory of explanation (see next chapter). I think that the causalist needs to enlarge their explanatory schemas. I address these issues in the order of their relevance to my case study.

In respect to (B) (see page 94), the winner-takes-it-all worry is central to my discussion because in Part III I compare unifications and explanations achieved at different stages of the same theory: the unification in Kaluza and Klein respectively; and explanation in Kaluza and in Klein respectively. *Pace* Woodward, I believe we can compare explanations before and after unification occurs. In his example, one can investigate whether

Ampere's law or Coulomb's law explain less than, say, the laws of the covariant **EM**. I believe local and limited comparison between explanation and unification are kosher. I prefer to analyze explanations provided by a theory T indexicalized to a context C. Do we want to think of explanatory power of a theory as a *global* feature of a theory? In fact it is clear that some explanations score worse than others in different contexts. In Woodward's interpretation, Kitcher's unificationism works like a necessary and sufficient condition to explanation. But both Kitcher and Friedman suggest that unification is context dependent. Unificatory elements are unificatory in a given context and not in others: for example the absolute reference frame unifies in Newtonian mechanics only. Take Maxwell's unification and put it in the context of, let us say, quantum electrodynamics: it does not unify anymore. Absolute spacetime does unify in the classical mechanics, but it plays no role in let us say, **GR**. Coulomb's law was explanatory in the absence of Maxwell's theory and became less explanatory once Maxwell's theory emerged. Of course, absolute comparisons are difficult, if not impossible, and they tell us nothing about how science in fact operates and uses unifications. Imagine that you want to compare the power of explanation of the evolutionary theory to the power of explanation of the Big-Bang cosmological model: This is totally hopeless. But local comparisons could work. My case study illustrates this: in Part III I show that comparing Kaluza's and Klein's unifications or explanations is possible and relevant.

But there is a serious problem here: if Woodward suggests that pre-Newtonian theories had explained better because they were prone to causal explanation whereas Newton's was not, causal explanation is in deep trouble.[56] If the choice is between:

- Newtonian physics: describes pretty accurately the phenomena, postulates gravitation and does not use hypotheses (does not causally explain)
- pre-Newtonian physics: explains causally the phenomena and uses causal hypotheses

what is the obvious choice? It seems that the causal explanation is not immune to crass pseudo-explanations. But it is not always the case that we can disqualify the explanatory theory and favor the descriptive one.[57]

In respect to (A) (see page 91) I agree that unification is a complex and diverse scientific achievement, but the categories described in Woodward are not exclusive: one can see unifications which are combinations of types (I)-(III) at different moments of time or even at the same time. For some reasons, Woodward ignores cases of classification in science which were explanatory without being causal. I mentioned Mendeleev's classification of chemical elements by their atomic mass. It explained the property of some metals by their place in the periodic table; it predicted the existence of unknown elements; it explained why some elements are mostly solid and why some others are gases. This classification in particular was unificatory in disguise. Half a century later on, once the quantum

---

[56] Ditto the pre-Maxwell theories which were based on the theory of the ether. We know now that Maxwell's theory admits an ether-free interpretation.

[57] Newtonians explained inertia by identifying it with a property of spacetime; Mach's explanation of inertial forces was in terms of inertia exerted by other bodies in space as a form of correlation. Both explained, although they had different predictions for the crucial experiment with two isolated rotating masses connected by a cable. This was an older debate between Sklar and Wolfson & Woodward from the mid 1970s For Woodward & Wolfson, Mach's explanation is causal (i.e. by using correlations), whereas Newton's is geometric (by identification). But Woodward & Wolfson took both as being immune to pseudo-explanations (Sklar 1974; Wolfson and Woodward 1979).

theory had been discovered, one could causally connect the property of elements to the presence of electrons and nucleons. Moreover, once the theory of nuclear reaction had been developed, even a manipulation of chemical property became possible.[58] Reclassification of optical phenomena after Maxwell's unification of optics and electromagnetism is one case. Also Newton's unification shed a new light on celestial phenomena and the classification of comets for example. There is a deeper problem here than the terminological one. If not all classifications are explanatory, then how do we discern those which are from those which are not? Woodward's answer is: exclusively by a theory of causation. This raises some concerns when one looks at the science as a whole because it is difficult to convince philosophers that the same theory of causation would work in all sciences.

Type (II) (see page 91) unification is central to my discussion. Is mathematics a necessary condition of unification? I answer this question in the negative. I do not suggest that (II) is the only unification possible, but I think that Woodward's argument that mathematics does not explain at all is misleading. It seems that Woodward thinks that mathematics is necessarily associated to unification and, as he excludes mathematical explanations from the genuine scientific explanations, *voilà*: unification does not explain.

There are two possible answers to this argument. First, one could say that mathematics explains. Geometry explains, solutions to **PDE** explain, you name it. Spacetime explains (Nerlich 1994, 283); group theory in Standard Model explains whether and why a specific elementary particle exists and why another one does not. All these mathematical structures are unificatory in some respect. So Woodward is wrong in dismissing mathematical ex-

[58] I suspect the same is true of Linnaeus' biological classification. Once we have the genes, the classification of animals can be linked causally to something real in the world and can be manipulated, too, within certain limits.

planations. This is the line of thought that applies to the present dissertation. But there is a second way to parry away Woodward's concern. Let us suppose for the sake of the argument that mathematics does not explain at all by itself, or at least in the absence of some causal hypotheses. But is it that unification always comes in a mathematical wrapping? I mentioned already that some scientists would say that mathematics is unreasonably effective in theoretical sciences. In what sense? In providing accurate descriptions and predictions. Sometimes physicists realized that the mathematical formalism had been developed a long time ago. As we will see, this was the case with **GR** and implicitly with Kaluza-Klein theories. As S. Weinberg puts it, it is "positively spooky how the physicist finds the mathematician has been there before him or her" (Weinberg 1986 722).

If mathematics is sometimes "unreasonably effective" in describing nature, is it effective in unifying theories, too? The fact that we employ the same mathematical formalism in two different theories is not unification yet. For example, the dynamical systems theory is extensively used in a vast array of disciplines: cognitive sciences, physics, economics, biology etc., and it has an incredible power of systematization, although it is not unificatory. These cases radically cross-cut over several areas and are able to provide the right equations for the dynamics of systems (Smith 1992). But they are not unifications. The answer is: because they do not explain and there is no trivial theory that puts together the two domains of phenomena. Postulating for example that there is a thing in the world called "chaos" that is instantiated in several systems, no matter what their material constituency is, is not an attractive way of unifying theories. But is there explanation besides causal explanations? As I discussed before, I claim that there are other types of explanations than causal explanations and I illustrate this in Part III of my dissertation. It is important to keep in mind that unification can occur when there is no accompanying mathematics and the other way around: mathematics does not entail unification.

In many cases, mathematics has not been there when unification occurred. Some examples are in order here. Major discoveries in science were deprived of mathematics: Newton needed to develop calculus in order to deal with the new mechanics. The initial development of quantum mechanics involved a complex procedure of rewriting the mathematics of non-commutative operators. In both cases mathematics was recreated or reinvented for specific purposes of physics. Think also of Darwinian unification.[59] Mathematics was not in the party.[60] A similar situation seems to occur in the current development of String Theory because, as many specialists claim, mathematics of String Theory is non-existent or underdeveloped, or in other words, String Theory is its own mathematics. Similar to these developments, unification can be present even where mathematics is absent or it has not come yet. In sum, some theories unify without being mathematically formalized, sometimes mathematics plays a secondary role within unification and some highly general theories with a powerful mathematics under their sleeve do not unify.

In short, I see the differences between (I), (II) and (III) (page 91) as contingent and as a matter of historical perspective. Sometimes and maybe in many cases of exemplar unifications one can see all three types of unification at work.

[59] There is an interesting discussion on the unification of Mendelian genetics and Darwinian evolution in the works of R. Fisher and S. Wright at the beginning of the 20th century in (Morrison 2000 ch. 7).

[60] One could say that Malthus' theory which is partially mathematical played an important role in the genesis of the evolutionary theory.

4.4. Interventions, modularity and unification

I do not have a specific point to make about "deductive chauvinism" (C). This problem does not affect my case study because Kaluza-Klein has a distinct deductive structure, difficult to dismiss (in the Part II, I schematize Kaluza's and Klein's arguments). I suggest that Woodward's criticism is "causal chauvinistic" because it excludes non-causal explanations. He thinks that in order to have scientific explanation we need causality, more precisely causality by interventions. I want to insist on this specific point in this section.

Let me sketch here Woodward's theory of causation. He distances himself from existing interventionist theories: Reichenbach's condition of "screening-off" (1956) and Menzies and Price's agent-based theory (1993)—an explicit anthropomorphic account—as well as Salmon's theory of causation (1984). For Woodward, intervention does not involve agents, or experiments, even possible in principle. His theory is based on counterfactuals and improves supposedly upon the causal Markov theory. Besides some advantages, one of the problems of the interventionist theory is that it is circular: the definition of intervention presupposes causation (Woodward 2003 98).

In this framework, we have a system of variables **V** that describe a system. What is an intervention? An intervention is a way of isolating the influence on a variable X_i; it cuts all the arrows directed into X_i except the intervening variable. Let us take $P_{[Z_i=on]}(X_j)$ the probability of X_j if the intervention variable Z_i is "on" and $P_{[Z_i=off]}(X_j)$ the probability of X_j if there is no intervention. Here is a quick definition of causation via "modularity":

Def 5 **MOD*:** For all X_i and X_j different variables in V, if $P_{[Z_i=on]}(X_j) \neq P_{[Z_i=off]}(X_j)$, then X_i causes X_j.

Informally, causation can be read as a counterfactual: an event C is the cause of an effect E if C were to be changed by an intervention, E would change (Woodward 2003 132; Hausman and Woodward 2004 851).

Woodward is quick in criticizing unification, but unification and his interventionism apparently have more things in common. His interventionism is in fact dependent on scientific theories and moreover on how unified the theories are. The key concept used by his theory is modularity and this is theory laden in the context of Kaluza-Klein theory. Let me give here a toy model of what I mean by modularity depending on scientific theories from Maxwell's unification. With some changes, this story can be replicated in the case of Kaluza-Klein theories.

Let us think that an interventionist dis-unifier decides to keep electricity and magnetism separated. We take a system **S** and describe it by some variables in **V**. For the sake of the arguments, let us presuppose that **S** is neither electrostatic, nor is it constituted by a magnet at rest: **S** could have currents, moving charges or anything more than objects at rest in it. If the dis-unifier is right, his description of the system is modularized into "electric variables" (X^{el}) and "magnetic variables" (X^{mag}). She thinks she can keep the two theories separate (let us symbolize this theory as **E+M**). She would simply deny that there is a causal relation between them by postulating a specific causality within **S**:

Def 6 **E+M causality**: For all X_j^{mag} and X_i^{el} variables in V, it is always the case that:

$$P_{[Z_i=on]}(X_j^{mag}) = P_{[Z_i=off]}(X_j^{mag})$$

$$P_{[Z_j=on]}(X_i^{el}) = P_{[Z_j=off]}(X_i^{el})$$

In other words, for the dis-unifier there are no electric interventions with magnetic consequences and there are no magnetic interventions with electric consequences. From the point of view of explanations, she claims that there are no electrical *explananda* with magnetic *explanans* and the other way around.[61] These claims of the dis-unifier are clearly false: we see magnetic effects with electric causes and the other way around. This means the modularity picked by this dis-unifier is not the right one.

A second dis-unifier comes and modularizes the system differently. Let us say we put a "sad person" and a "happy person" in the same room with system **S**. This dis-unifier accepts that there is a electromagnetic theory (**EM**), but rejects an electromagnetic-psychic theory. Then she modularizes the variables in "electro-magnetic" X^{em}, and "moods" X^{mood}. She thinks that there are no "moody" interventions Z^{mood} that have electromagnetic consequences and that there are no electromagnetic interventions with psychic consequences (in some certain limits, of course).

Def 7 **EM+PSYCH:** For all X_i^{em} and X_j^{mood} different variables in **V**, then

$$P_{[Z_i^{mood}=on]}(X_j^{em}) = P_{[Z_i^{mood}=off]}(X_j^{em}) \text{ and } P_{[Z_i=on]}(X_j^{mood}) = P_{[Z_i=off]}(X_j^{mood}).$$

This second dis-unifier might be right from an empirical point of view.

My suggestion is that unification is a way of picking the modularity because modularity is "theory-laden" and it depends heavily on the theory we adopt. A good unification will pick the right modularization. Another unificationist, the one who believes in

[61] For the moment I do not dispute the empirical falsity of all these claims.

T. More's "spippisitude" (a moral dimension of the world), would pick another modularization.[62] Another unificationist who believes in a **TOE** will pick again the wrong modularity and consequently the wrong explanations. Also, an extreme dis-unifier would excessively modularize the domain of the variable and will end up with the pseudo-explanations. The point is that extremists, both the dis-unifier and the unifiers, will pick the wrong modularization. In a nutshell, Woodward's own theory is a suitable path to discuss the way in which good explanations are generated from the right unificatory theories via the concept of modularity.

4.5.Pluralism about explanation and causation

What I suggest here is that: (a) there are several types of explanations other than the causal explanations and (b) the rift between causal explanations and explanations *qua* unifications can be bridged in several relevant ways.

This is a strategy followed by both camps. W. Salmon, once the leading figure in the ontic approach to explanation, admitted that at the end of the day unification is a key part of the explanation. In later work he hinted toward the idea that unification and causation are two sides of the same coin (Salmon 2006, 234; Salmon 1998, 434). Similarly, Kitcher seemed more and more convinced by recent advancement in the theory of causation that it was not so dubious a concept after all.[63] Seemingly, by the early 2000s, both authors were very akin to a pluralistic or at least dualistic approach to explanation in which both causation and unification may or may not work together. According to Peter Godfrey-Smith, "we cannot get the right analysis by claiming that within [all of science], a good

[62] I discuss More's theory in Section 15.6.
[63] This is anecdotally reported by P. Godfrey-Smith.

explanation is something that satisfies *either* the causal test *or* the unification test (etc.)" (Godfrey-Smith 2003 197). For Godfrey-Smith, this leaves out the way that different scientific fields will establish definite criteria for what is a good explanation. He advocates "contextualism" in respect of explanation: the standards of good explanations depend partially on the scientific context.

I have to add that Godfrey-Smith's contextualism, very akin to van Fraassen's account of pragmatic account of explanation, faces the problem of trivialization.[64] Change the context or be lenient enough with your context, and anything can count as explanation. How do we discern bad, trivial, spurious, or unnatural explanations from the good one? The same question applies here: how do we take apart the chaff from the wheat? As Godfrey-Smith rejects the orthodox approach of the "inference to the best explanation", I see here a problem with this form of unqualified pluralism. The question is to get a decent form of pluralism without blowing it into relativism. Although I embrace pluralism with respect to explanation, one can see how for specific theories and contexts unification bolsters specific types of explanations. But I avoid entering this debate at the general level. In the context of my case study, I analyze the active role unification plays in construing explanations.

One can argue here against my proposal by pointing out that because unification itself can be trivial/spurious/unnatural etc., the issue cannot be settled that easy. One may need to go back to causation and laws of nature. Some philosophers, commonly labeled "Humeans", tried to explain laws of nature and causation in terms of patterns in the

[64] Maybe T. Kuhn's approach to explanation is actually close enough to this type of contextualism.

structure of the world and shunned away from the causal talk.[65] Unsurprisingly enough, unificationism is fond of this approach in its reluctance to talk about causation.

My short reply (details to be developed in Part III) is that in the early episode of the Kaluza-Klein theory, causation was not directly present and the wholehearted unificationist would find satisfaction in the way in which unification produced novel, unintended explanations. The Kaluza-Klein explanation is not causal. Causation is present in Kaluza-Klein theory as it is present in **GR** and **EM**. But a Kaluza-Klein theory in **5-D** does not explain why billiard balls collide, why I can cure fever with an aspirin, etc. More precisely, it does not do it in virtue of being unificatory theory, or alternatively, if it does explain such facts it does it as **EM** or **GR** do. I provide a partial answer at best to the thorny issues of unification as an alternative to causal explanation by showing that in the case of Kaluza-Klein theory, causal explanations come after non-causal explanations are established. In other words, causal explanation did not play any role in the original formulation of the theory, although later they be would eventually added. There is an interesting story to be told about causation in *later* developments of the Kaluza-Klein story in which causation is involved in the explanation undertaking of the theory. Those things to come in Part III.

[65] See also recent work of Callender and Cohen in which they use a form of contextualism in the Mill-Ramsey-Lewis approach to laws of nature (Callender and Cohen 2009 forthcoming).

Chapter 5. A deflationary account of unification (M. Morrison)

In direct relation to Godfrey-Smith's contextualism and Woodward's criticisms of unification, I discuss Margaret Morrison's deflationary account of unification.[66] From my point of view, Morrison's work is the closest to my approach because she analyzes several cases of unification and criticizes both formal and general approaches to unification. For Morrison, unification needs to be analyzed on a case by case basis and not through a general account like Kitcher's, Friedman's, Glymour's, etc. I call her approach the "deflationary account" as she expresses several times her skepticism against any formal or general accounts of unification and paints a rather negative picture of unification in science. Morrison provides the major motivation for her analyses:[67]

> I have […] suggested that unification is not the kind of criterion on which to base arguments for realism, and I have also hinted that it may not be important for theory acceptance either. In order to substantiate that argument with empirical evidence, I want to examine some specific and paradigmatic cases of theory unification in both the physical and biological sciences. One of my claims is that unification typically was not considered to be a crucial methodological factor in either the development or confirmation of the physical theories. And even in cases where it was a motivating factor, […] the kind of unity that was produced could not be identified with the theory's ability to explain specific phenomena. […] I want to demonstrate that the ways in which theory unification takes place and the role it plays in scientific contexts have little to do with how it has been characterized in traditional philosophical debates. Once we have a clearer understanding of the unifying process we can begin to see where its importance lies, what its connection is, if any, to explanation and the way unity functions in partic-

---

66 See especially *Unifying Scientific Theories: Physical Concepts and Mathematical Structures* (Morrison 2000 59). Occasionally, I will refer to her previous work (Morrison 1995; Morrison 1990; Morrison 1992) or to her more recent controversy with A. Plutynski on explanation and unification in theoretical biology (Plutynski 2005; Morrison 2006).

67 The chapters containing her case studies are: Ch. 3: *Maxwell's Unification of Electromagnetism and Optics* ; Ch. 4: *Gauges, Symmetries and Forces: The Electroweak Unification*; Ch. 5: *Special Relativity and the Unity of Physics*; Ch. 6: *Darwin and Natural Selection; Unification versus Explanation*; Ch. 7: *Structural Unity and the Biological Synthesis*; each corresponding to a separate unificatory theory.

> ular domains as well as in the broader context of scientific inquiry (Morrison 2000 59).

There are several strong claims here. First, in the cases she studies, she finds that either unification is not a motivating factor, or where it is, it is not identified with the theory's ability to *explain* specific phenomena. Second, she finds no reasons to link unification to scientific realism. Third, unificatory theories do not play a significant role in confirmation of empirical data. Last but not least, these theories do not seem important from a methodological point of view. All three claims are justified by analyzing several instances of unification.

Although Morrison does not provide a general "theory of unification"—as she claims that no such account is possible—there nevertheless seem to be "good reasons for thinking that theory unification is more clear-cut" than the "unity of science" doctrine which has several interpretations (Morrison 2000 29). Morrison does not argue against the dis-unity of science movement by providing counterexamples of unity in science, but by critically analyzing the cases of unified theories and unifications. The deflationary account is more or less a descriptive one: she draws attention to some general features of unified theories and their philosophical consequences. In order to reject Kitcer-Friedman-Glymour accounts of unification that relate it to explanation or realism, Morrison cashes out some features of unified theories and describes the unificatory process as well.

In one sense her analysis is more modest, but in another sense more ambitious. It is true that she does not need to rely on a general theory of unification. Unification is a multi-faceted process that involves many components besides explanation and understanding; as Woodward would suggest, it is heterogeneous. A description of each separate case would suffice. But from the beginning there is room for an ambiguity in Morrison's argument. A natural question is: how broad are Morrison's conclusions?

In order to relate Morrison's approach to my case study, I further interpret her claims by quantifying them. We have a certain number of unifications in science; only some of them are successful and even fewer are "exemplar" or perfect.[68] Morrison analyses some of the successful and exemplar unifications and draws the conclusions mentioned above. But her claims extend beyond the cases studied. Here are some relevant quotations: in the "Introduction", when she states for the first time that unification and explanation are decoupled she says that "rather than analysing unification as a special case of explanatory power, as is commonly done in the literature, I claim that they [unification and explanation] frequently have little to do with each other and in many cases are actually at odds." (Morrison 2000 2) In arguing against Kitcher, Morrison shows that explanation and unification cannot be related in general: "I want to argue, using special examples of unified theories, that the mechanisms crucial to the unifying process often supply little or no theoretical explanation of the physical dynamics of the unified theory." (Morrison 2000 4).

My interpretation to Morrison's project is that she suggests an extension of her analysis to other cases of unification not discussed in the book by using quantifiers such as "typically", "frequently", "in many cases". There is another reason to think that she envisages such a generalization to other cases: in criticizing Friedman (1983) she dismisses the connection between unification and realism. But Friedman's analysis is couched in terms of **GR**, a theory not discussed in Morrison's book, at least not as an un-

---

[68] I discuss T. Maudlin's more sophisticated ranking of unification in Section 12.6.

ificatory theory. If Friedman's analysis is wrong, then Morrison's conclusions can be extended to **GR** as unificatory theory.

Here are the most important claims of her book. Although she lists them at the beginning, I state them in the quantified form suggested before (the bold text is the "multal" quantifier added by me):

> ***[5]*** *Unification and explanatory power are decoupled—in the vast majority of unifications, exemplar or not.*
>
> ***[6]*** *Unification presupposes a mathematical structure or mathematization of the phenomena and a concept or a parameter* ***—in the vast majority of unifications, exemplar or not****.*
>
> ***[7]*** *Unification has no metaphysical or ontological implications for the "unity" in nature* ***—in the vast majority of unifications, exemplar or not****.*

My main goal is to argue against [5] and [6], i.e. the quantified versions of Morrison's claims, but incidentally I argue against some of her interpretations to Maxwell's unification and to the unification in **SR** (mainly in Part III).

In order to provide a cogent argument in favor of [5], Morrison questions the very core of Friedman's supposition: does a theory unify a group of phenomena? It is a fact that Newton's theory unified celestial and terrestrial phenomena. Nowadays, we no longer accept the dynamics required to make that theory explanatory. By separating explanation and unification we entertain our intuitions about independence of theory unification while recognizing the historical aspects of explanation. In other words, we still accept the unificatory power of Newton's theory, but we admit that his explanation is dated, thus we "decouple" the explanation from the unification. The main aim of Morrison's book is to determine the extent to which a particular theory has unified different domains and to show that the explanation provided in each case is at best precarious.

Why does Morrison reject the connection between unification and explanation? Because for her unity is understood in terms of "derivations" and derivation does not explain how processes are unified:

> […] explanation and unification may not be as closely related as has typically been thought; unity is possible without a satisfactory level of explanatory power. Moreover, when unification is analyzed in terms of something like the **D-N** framework, it becomes clear that the account of unification that results provides virtually no understanding of *how* the unifying process takes place. Because unity is understood simply in terms of derivability, there is no sense of how the phenomena become integrated within a theoretical edifice (Morrison 2000 4).

In other words, Morrison is closer to Woodward's criticism against Kitcher's "deductive chauvinism": if unification is a deductive process, explanation need not to be like that. For Morrison, there is something that the unificationist is distorting: causation. Her rejection is rooted not in the case by case analysis of unification, but in the account of explanation that she assumes. In other words, her analyses are deeply centered on what she means by explanation and less on unification itself. I believe she takes the side of the causal explanation discussed in the previous chapters and she dismisses explanations that are not causal.

In respect to [6] (p. 111), Morrison identifies in all cases a theoretical structure that either represents or facilitates the unifying mechanism. But, similar to what Woodward suggested, it has no explanatory power:

> To see how this is so in the context of Maxwell's theory, we need only look at the way the Lagrangian formalism functioned in allowing Maxwell to provide a dynamical theory without any explanation of the physical causes

> that underlay the phenomena. The generality of the Lagrangian approach makes it applicable in a variety of contexts, and it is ultimately this nature that makes it especially suited to unifying different domains. But this generality has a drawback. By not providing an account of the way physical processes take place, the unifying power is achieved at the expense of explanatory power (Morrison 2000 64).

Using examples of unified theories, Morrison argues that the mark of a *truly* unified theory is "a specific mechanism or theoretical quantity/parameter that is not present in a simple conjunction, a parameter that represents the theory's ability to reduce, identify or synthesize two or more processes within the confines of a single theoretical framework" (Morrison 2000 64). By attacking Glymour's thesis that theories can eliminate contingencies by necessity, Morrison claims that using identities between a mathematical structure or parameter and a real process in the world as a tool to eliminate contingencies can in fact have opposite effects. Identifications are in many cases not unique and different interpretations will use different identities. When they are absent (for Morrison this is the case of **SR**, electroweak theory and biological synthesis) we simply have unity without explanation:

> Often an identification of a phenomenon with a particular mathematical characterization is highly contingent, and the generality of such frameworks is such that they provide no unique or detailed understanding of the physical systems that they represent. We […] predict the motions of phenomena from dynamical principles, but we have no understanding of the causes of motion. Hence, there is no guarantee of explanatory power resulting from the mathematical description afforded by our theories (Morrison 2000 31).

5.1. Two types of unity

In arguing for [7] (p. 111), Morrison draws a distinction between synthetic and reductive unity. Reductive unity presupposes that if two phenomena are identified as being of the same kind, one of them can be eliminated (Morrison 2000 section 2.3). Celestial phenomena are governed by the same laws as terrestrial one; *ergo* celestial mechanics and terrestrial mechanics are identified under the same formalism. Terrestrial description has some practical advantages such as empirical accessibility and easy manipulability over the celestial. For all practical purposes, even the formalism can be different. Physicists still talk about astronomical objects and small size objects, but they use the same representation and laws to describe them. In the case of caloric theory, its vocabulary has been completely reduced to the language of heat theory and consequently caloric theory disappeared because of its major problems. It is not needed anymore, so it is replaced by thermodynamic descriptions. Thermodynamic phenomena can be described by statistical mechanics and all thermodynamic variables can be deduced from statistical variables of the system. The same can be said about Maxwell's reduction of optics to electromagnetism.[69] According to Morrison, the connection between electromagnetism and optics was, however, a reductive one. Light and waves both travel in one and the same medium with the same speed. Hence optics is redundant at a fundamental level. The luminiferous ether has simply disappeared from the theory, being replaced by the electromagnetic ether. Maxwell successfully eliminated any references to optics and light from his theory, ontologically and epistemologically reducing all objects and laws of optics to electromagnetic objects and laws. The electric-optic duality ceased to exist in Maxwell's theory. In other words, reductive unification can remove a duality and can totally eliminate the language of a theory. Nicholas Maxwell called the reductive unification of optics and electromagnetism, "unification by annihilation" because any spooky property of light is revealed as a property of electromagnetic waves having a specific wavelength (Maxwell 1998 125-6).

---

[69] More details on the reductive unity can be found in (Bechtel and Hamilton 2007).

Here is a problem: fewer and fewer philosophers are convinced that we have genuine reductive unifications in science.[70] Hard-core reductionism barely survived the halcyon decade of the **D-N** models, although there is a new wave of reductionism (see Section 12.4). Unification plays a special role in the attempt to replace scientific reduction. Where there is no reduction, unification can play a significant role. In general, theoretical unification does not entail theoretical reduction. Morrison admits that some theoretical unifications are not reductive, but rather *synthetic unifications* (Morrison 2000 107)**.** They involve the *integration* of two separate processes or phenomena under one theory (Morrison 2000 5). The reduction and the deduction of one theory to/from another do not work even in simple cases such as the Newtonian theory unifying Galileo's laws with Kepler's laws because Newton predicted different trajectories for falling bodies and satellites than Galileo and Kepler. Similar things can be said about the reductive unification of thermodynamics to statistical mechanics which does not work seamlessly, *pace* Friedman (Friedman 1983 254sqq.; Morrison 2000 sect. 2.3).

For my present purposes, the synthetic unity is more interesting and more recurrent. As exemplar unification, Maxwell's **EM** theory is a synthetic unification. In comparing electromagnetic unification to electroweak unification, Morrison noticed that:

> a reconceptualization of the electromagnetic potential and a new dynamics emerged from the mixing of the fields. […] the phenomena are interpreted in a new way, an interpretation that results not from conjoining two theories but from a genuine synthesis—in this case, a synthesis that retains an element of independence for each domain but yields a broader theoretical

[70] The thermodynamics-to-statistical mechanics case of reduction or the caloric theory-to-thermodynamics case or even the Newtonian case can be questioned on various reasons. Arguments against reductive unification are not arguments against unification in general. For an anti-reductionistic approach to optics for example, see (Batterman 2002 ch. 6).

> framework within which their integration can be achieved (Morrison 2000 34).

The equivalence of mass and energy is another example of synthetic unity. Synthetic unification "integrates" two phenomena in one mathematical framework. On the other hand, within the reductive unity, two phenomena are identified as being of the same kind (such as optical phenomena are electromagnetic phenomena in Maxwell's unification). Morrison suggests that only this type of unity could have ontological implications, even if they are not straightforward. Even if it is not explicitly, Morrison thinks that reductive unification is the most worthy because all aspects of unification are fulfilled within reductive cases of unification. I do not see how Morrison can address the serious problems that reduction faces today. Criticisms against reductionism affect indirectly reductive unification. Secondly, there are intermediate cases between synthetic unification and reductive unifications rarely discussed. In my case study I show that non-reductive unifications, synthetic unifications included, are interesting and relevant to the scientific progress, too.

I think that the synthetic unity is more interesting and more recurrent from a philosophical point of view. Maxwell's **EM** theory is a synthetic unification. Another example of synthetic unity is the equivalence of mass and energy in **SR**—for Morrison a case of synthetic unification which "integrates" two phenomena in one and the mathematical framework. Moreover, as a general point, I claim in Part III that non-reductive unifications are more powerful than reductive ones in respect of explanation.

Claim [7] (p. 111) is a direct attack against Friedman's "realism as unification" thesis. Morrison's attack is twofold:

a) *The model/sub-model relation and unification.* Morrison does not endorse Friedman's relation between model and sub-model. If the observational structure is a subpart of the theoretical one, then Friedman's account cannot render the evolution of a theory as a change between observational and theoretical structure. Morrison thinks that the model/sub-model approach is not capable of characterizing scientific theories. Moreover, the identity relation postulated by Friedman is too tight to allow that "kind of looseness to fit that exists between the theoretical and observable structures of theories" (Morrison 2000 38).

b) *Realism and unification*. For Morrison, the main problem with Friedman (1983) is his emphasis on the connection between realism and unification. Morrison holds that Friedman claims that in order to be realist, one needs unification. The natural question Morrison asks is: what kind of realism does Friedman need? Friedman is not able to make an elementary distinction between two kinds of realism, i.e. semantic realism and epistemological realism. As Friedman interprets the theoretical structure literally, his semantic realism enables us to give derivations and entailments of phenomenological laws without relying on phenomenological properties. For Friedman, successful conjunction provides evidence for truth and through conjunction theories pick up boosts in confirmation. This does not constitute a reason to believe that the theoretical structure is real. It could be real, but there is no commitment either way. But conjunction requires epistemological realism , i.e. the position that theoretical hypotheses are trueOnce we realize that science operates with more than truth-functional operators, we need to move beyond the mere conjunction.

One can take Friedman only as a semantic realist without epistemological consequences. In this case, Friedman's position is not anymore tenable. In this case he gets too close to van Fraassen, whose position Friedman criticized vehemently. By making the distinction between a logical and a methodological conjunctive inference, Morrison shows that neither of them is appropriate to explicate unification.

In sum, for Morrison, realism cannot be bolstered by Friedman's theory of unification. Moreover, Friedman's position commits the same sins as the original formulation due to Boyd or Putnam (Putnam 1979, 457). She insists that the literal interpretation of theoretical terms can bring up troubles. Her example comes from an exemplar case of reduction, the reduction of thermodynamics to statistical mechanics. In phase space, each velocity has three components on the three axes: v_x,v_y, v_z. Morrison asks if one can identify an amount of energy to all these components, i.e. if there is a real energy having the value $\frac{mv_x^2}{2}$ for example. By way of this quick example, she claims that there is no identification of energy with a mathematical representation. I cannot see here the relevance of her short argument: there are indefinitely many components of a real entity that have no physical meaning. Other components do have explanatory power and can make predictions even if they are not directly identifiable with physical objects. Take for example phase space in thermodynamics. Although she admits that the mathematical representation in phase space is crucial for modeling statistical systems: "even if we disregard the problem of identifying temperature and mean kinetic energy across theoretical boundaries, a more significant difficulty arises in the case of *identifying*, in the way suggested by the model/sub-model approach, the constituents of the system postulated by classical statistical mechanics with individual particles" (Morrison 2000 46). Friedman also acknowledges that if one is interested in pure phenomenological laws, one can use representation instead of identity.

Nevertheless, if one aims to greater confirmation of hypotheses, one needs the literal identification. Even in exemplar cases as the reduction of thermodynamics to statistical mechanics, the literal interpretation prohibits statistical mechanics from accounting for specific parameters. The representationalist account can counter this difficulty. The apparatus of formal identification is simply too rigid to capture the complexity of the relation between model and phenomena.

For Morrison, the virtue of unification is neither empirical nor metaphysical, but epistemic. She states that a concept can be unificatory in a specific context in the same way as information can be useful in one context and not relevant in another one. So if a piece of theoretical structure is unificatory in such and such context but it is not in another one, than a realist would have hard times in using it as evidence for the truth of theories. Historical contingencies dictate the unificatory power of a concept. But one cannot infer from this any ontological claim. Contra Friedman, Morrison's conclusion is that conjunction and reduction do not play a role in unification.

5.2. Is causation the missing "machinery"?

Morrison thinks that explanations have to be related to the "machinery", i.e. the causal story of the phenomena described.[71] Only the "machinery" provides an answer to the "how" questions on top of the "why" questions proliferated by the **D-N** mantra. Moreover, the machinery gives us "understanding". The "machinery" is a fuzzy concept in

[71] Morrison mentions Feynman's name without citing or quoting his work. I tried to find a precise reference to Feynman and there are few options available. Feynman spoke incidentally about computers as "machineries", i.e. computational devices when he foresaw the development of nanotechnologies in the 1950s. Others refer to "Feynman's machinery" as a method of computation scattering amplitudes in QFT. All these meanings are remote from what Morrison suggests. The suggested meaning is machinery as causal and mechanistic explanation. I plan to contact her for further clarifications.

Morrison and I suppose she refers loosely to something at the border between "causation" and "mechanical explanations", maybe a "causal-mechanical explanations". Some quotations are in order here:

> Explanation by derivation from quantitative laws very often doesn't provide what Richard Feynman calls the 'machinery' of a particular system. The machinery is what gives us the mechanism that explains why, but more importantly *how* a certain process takes place. When we ask about the propagation of electromagnetic waves, we want to know not just their velocity but also *how* they travel through space and the mechanism responsible for their propagation. Maxwell's first account of electrodynamics explained this in terms of the ether (Morrison 2000 3-4 my emphasis).

But as Morrison acknowledges, even the paradigmatic example of Newton's mechanics, unlike Descartes' physics, misses the "*how*" question entirely:

> […] one of the most striking features of the *Principia* is its move away from explanations of planetary motions in terms of mechanical causes. Instead, the mathematical form of force is highlighted; the planetary ellipses discovered by Kepler are 'explained' in terms of a mathematical description of the force that produces those motions. […] there is no explanation of how this gravitational force acts on bodies (how it is transported), nor is there any account of its causal properties" (Morrison 2000 4)**.**

Originally, Maxwell explained the propagation of the electromagnetic wave by employing a machinery called the "vortex ether model" (Morrison 2000 71; Maxwell (Clerk) 1861). Later on, Maxwell shifted towards a dynamical theory of electromagnetism that was about the field in the space around electric and magnetic bodies and not about the ether. Long before Maxwell's unification, J. MacCullagh and G. Green found a potential function for the ether without postulating its mechanical features and hence the equation of the **EM** wave was inferred without specifying the details of the mechanical structure of the ether. Ether was treated as a mechanical system "without any specification of the machinery that gave rise to the characteristics exhibited by the potential-energy function"

(Morrison 2000 82; Maxwell 1865 1:564). Similar to Newton's celebrated phrase *hypotheses non fingo*, Maxwell explicitly said that he "avoided any hypothesis of that kind" and added that terms such as electric momentum, electric elasticity are only illustrative, not explanatory.

In the case of Darwin's theory of evolution, Morrison claims that the unificatory element was not the natural selection per se, as the result of Malthus' theory, but as a *vera causa* in Whewell's sense. Malthus helped Darwin understand how selection could be applied to organisms. "The missing link was a causal mechanism for understanding the selective process" (Morrison 2000 205). Malthus' law played a quantitative role; but it needed a causal interpretation that came only with Darwin. In this specific context, Morrison asks a very important question: does unity produce explanation? (Morrison 2000 section 6.4) Does a more unified theory provide us understanding? Natural selection in itself is not explanatory, it could function in an explanatory way only in conjunction with other assumptions and conditions that are strongly related to the causal models of evolution. Moreover, in later debates with Mendelians, Darwinians pointed to selection as the cause of evolution as opposed to mutation. Darwin's theory is both explanatory and unificatory, but the two are decoupled because what unifies does not explain and what explains does not unify.

I want to go back to the machinery mentioned by Morrison in the "Introduction". What is the machinery? The dynamic explanations, one that can be found in Newton, in the later development of analytical mechanics, or in the later work of Maxwell, gives some information and provides predictions, but "we frequently want to know more than that; we want to know about the machinery, part of which involves knowledge of the causal behavior of the system. It is this feature that enables us to understand how certain processes take place." (Morrison 2000 3-4). The "machinery" needs the causal structure of the world among other things. In some cases, Morrison admits, "such an explanation is not always possible; we may simply be unable to determine the material behavior or conditions that produce a particular event or effect. But even when this type of explanation is not available, the theory in question still may be able to unify a group of phenomena" (*idem*).

5.3. Morrison's argument revisited

Despite the breadth and strength of her analysis, Morrison is not analytically clear on unification and explanation. A definition of unification is difficult to concoct and maybe not that useful. Morrison rejects some possible candidates and argues against Friedman's embedding model of unification. The characterizations provided are usually vague. The same about explanation: she acknowledges that she is not in the position to offer a theory of explanation, except in criticizing Kitcher's. Much of the critical reactions to her book and her case studies (I count here M. Steiner, T. Jones, M. Liston, A. Plutinsky) have complained that Morrison never says exactly what she means by explanation and by unification (Muller 2001, 132-143). While I cannot go through all her case studies here, this remark is relevant here for the following reason: there are hints and suggestions throughout the book, where she is making claims about explanation, that she has *only* causal explanation or mechanistic explanation in mind. Remember that her approach is explicitly epistemological. I could not find a better equivalence for the machinery than the knowledge of the causal structure and I conclude that Morrison adopts the causal approach to explanation. In this respect, her position is on par with Woodward's and in conflict with Kitcher's "explanation as unification" approach. But what is missing in Morrison is the precision in characterizing both explanation and causation. Creating a middle term that loosely relates causation to explanation, i.e. the "machinery" does not help too much. But in her analysis of unification, causation plays a central role especially when we try to understand the "machinery" of unification. This is again controversial because her point of view contradicts what I called the pluralist approach to explanation. Morrison's argument is flawed because she rejects as genuine explanations other than causal ones. Seemingly, her account of explanation is not pluralistic because she minimize the importance of non-causal explanation. I adopt a more pluralistic point of view similar to Kitcher's, in which other types of explanations exist: functional explanations, structural explanations, etc. (see Chapter 16).

It seems that we have to add some hidden assumptions to Morrison's book in order to understand her approach:

[8] *Knowledge about causal behavior is a necessary component of the "machinery";*

[9] *Understanding of "the machinery" is a necessary condition for explanation;*

[10] *Unification does not need knowledge of the causal behavior and can be performed at the level of mathematical formalisms;*

[11] *In all the cases under scrutiny (i.e. in the vast majority of unifications), the progressive development of the mathematical formalism has been detrimental to the knowledge of the causal behavior of the system;*

> ***[12]*** *Therefore [5], i.e unification is present in the absence of explanation (in the vast majority of cases, exemplar or not).*[72]

A pluralist in respect of explanation would deny [9]. My claim is that in my case study there are other types of explanations that can function together and within unification. I do not plan to show that [11] is not true for the Kaluza and Klein theory, although I show that at a later stage of its evolution there is a causal interpretation of the theory that is not infringed by the mechanism of unification. Therefore, the machinery is one path among others to ascertain explanation.

Another problem is embedded in the conditional in [9]. Notwithstanding the subjectivity of understanding, its dependence on a set of previous beliefs and significations of observations, as Friedman warned us, let us take understanding as part of our epistemic goal—which I take to be explanation. Let us suppose that we manage to understand the machinery and we have a good grasp of it. As the history and the practice of science have witnessed, there are inferences from false explanations even when we have "the machinery". Ether, phlogiston, and a host of such "machineries" that worked well on the explanatory level as well on the unificatory level were in fact abandoned during the evolution of sciences because they did not exist. In other words, abductive inferences are not sound and there are such things as bad explanations even in the presence of the machinery.

If I understand Morrison's suggestion correctly, [10] is the claim that in moving towards more formal treatments, we lose the understanding of the machinery. The more detailed theories have more machinery under their sleeves. But in the case of Maxwell's

---

[72] I tried to render Morrison's argument in the simplest form using only hypothetical deductive sentences and *modus tollens*. In some specific cases (Darwin's theory, at least), things are more complicated and cannot be rendered in a sentential logic.

theory the machinery was the ether and its vortex model. The pre-Newtonian physics had a Cartesian model of the vortex. They had more machinery, but there were false. Do we want those types of machineries? Morrison's interesting suggestion is that in the cases discussed by her the unification was accompanied by a regress in answering "how" and "why" questions (hypotheses for Newton and Maxwell). Aristotelian physics was saturated with answers to "why" and "how" questions. Do we want them back? My suspicion is that even if Morrison is right about losing the machinery, in many cases that was beneficial and that the machinery harmed more than it helped. She cannot have her cake and eat it too.

Chapter 6. Why a philosophical appraisal of Kaluza and Klein?

6.1. Two approaches to unification

In order to better situate my contribution to the literature on unification from a methodological point of view, I systematize here the positions discussed so far in order to to put my own contribution on the map of the existing philosophical literature.

Formal (or general) approaches to unification. First, there is the approach to unification attempting to provide a general theory of unification that is expected to work both as a description of known cases and as a norm for possible future unifications. The best authors who illustrate it are M. Friedman, P. Kitcher, C. Glymour, W. Myrvold, J. Watkins N. Maxwell (Glymour 1980; to the existing references, add Schupbach 2005, 594-607; Lange 2004, 205-215; Myrvold 2003, 399-423; Maxwell 2004). Once it is presented and described in detail, a discussion of few examples follows. Maybe in an ideal world, a brilliant philosopher of science comes up with *the* account of scientific unification, independent of its instances and independent of other theoretical virtues. In this approach, the perfect unification encompasses all possible unifications ever, draws the distinction between bad unifications and the good ones and provides *the* canon for all possible future unifications. Any discussion of cases of unification are rendered either superfluous, or merely bolstering this perfect account of unification.

Things are much less than ideal. In the absence of a “Plato of unification” one needs to proceed in a more *a posteriori* manner. One reason is that unification is too vague and undetermined a philosophical concept, in the sense that there is no general definition or criterion available. For me a general approach sounds a little bit too bold. How could one deal with all cases of unification at once? Philosophers should acknowledge the diversity of cases of unification and that unification operates on very different types of phenomena. First, scientific unification is a complex enterprise and one cannot expect to capture it within one formal theory. Second, the content of scientific theory is essential and any attempt to formalize it or abstract away from it is fatal. Whether cheap or good, unification is likely to hang on alternatives available. It is likely to be context-sensitive. Is being a good unification or a bad unification linked to explanatory power? Arguably, yes—to some extent—and to some extent to other non-empirical virtues as well. Hence there will be many moving parts that depend on the context. Third, coming up with necessary and sufficient conditions for any theoretical concept is a complicated and rarely successful enterprise.

Hence I have reasons to endorse no formal approaches to unification *à la* Watkins, Friedman, Maxwell or Kitcher. Even Kitcher's approach, the most complex general approach available, is difficult to adapt to concrete cases.[73] My example does not literally instantiate universal schemes of unification similar to those proposed, if such a universal scheme exists, although I show that some aspects of Kitcher's formal theory of unification are easy to be identified. I do not think Kitcher's schema, advocated in 1981, is literally instantiated in the other cases of unification, even in cases explicitly discussed by him. It is useful to here draw a parallel between unification and induction because formal accounts of unification have a similar fate as formal theories of induction. Formal theories of induction were proposed and later rejected on similar grounds. John Norton showed in what

[73] Besides Morrison and Woodward, several authors criticized Kitcher by pointing out that his theory cannot account for induction (Barnes 1992; Jones 1995b; Halonen and Hintikka 1999; Weber 2002).

respect, unlike deduction, induction is local and specific.[74] As Norton claims that induction derives its license from facts, I claim that facts and specific contexts, as well as scientific practice give us the whole story of unification. In Norton's usage, all induction is *local*. I see a parallel here with unification. The tension between universality and the concrete success of each case of unification cannot be solved in general. I do not want to take this "locality" of unification at the extreme. Several interesting aspects of Kitcher's and Friedman's accounts deserve attention. Although I do not want to stress the deductive/formal part of Kaluza-Klein unification, I show in what respect it endorses Kitcher's and Friedman's formal "accounts".[75] Despite its popularity in the 19th century and the resonance the term "Consilience of Inductions" still has, the parallel between unification and induction is very limited in scope and plays only a heuristic role in the present analysis. I acknowledge that unification depends on background beliefs and unification cannot be reduced to universal schemas as it is closer to material facts than some formalists thought. I prefer not to fish in troubled waters and enter the debates surrounding induction. In short, unification is heterogeneous enough to not be approached at a formal level and both "unity" and "unification" have multiple meanings in both philosophy and in sciences.

Several cases of unification are vulnerable to charges of triviality, spuriousness or *adhocness*. According to some moderate enthusiasts, Kitcher included, unification is something we would like to obtain, but we hardly get there in the real practice of science. Very important questions such as: Is unification the norm or only a contingent feature of

---

[74] Mainly in his (Norton 2003), but also at a recent talk at the 21st Philosophy of Science Association Biennial Meeting in Pittsburgh, November 8, 2008. Formal approaches to induction were attempted by Aristotle, Bacon, Whewell, Mill, Russell, Carnap, Popper, *i.a.;* evidently, the number of those who tried to tackle a formal account of unification is much smaller. See some references at the beginning of this chapter.

[75] Here I simply refer to "Kaluza-Klein theory" as a whole. In subsequent chapters I will separately analyze them and I refer to Kaluza's or to Klein's theories.

some scientific theories? Is unification a consequence of the way we represent the world or does it say something about the world itself? have to be answered on a case-by-case basis. If the moderate enthusiasts are right, then unification can do this and that in specific cases, and cannot do this and that in other cases.

Witness the standard underdetermination argument for theoretical virtues of theories:[76] as two theories could have exactly the same empirical consequences, we need to invoke some theoretical virtues in deciding which theory to choose among various candidates.[77] As Kaluza-Klein is a theory in five dimensions, one can imagine that for any theory T, let us say **GR**, one can generate an infinite number of alternative theories T_1, T_2, T_3, etc. such that for example T_1= "there are five dimensions of spacetime, but all physical fields are null in the fifth dimension and there is no radiation of energy in extra dimension (and maybe other auxiliary restrictive conditions)", T_2 = "there are six dimensions of spacetime, but all physical fields are null in the fifth and sixth dimensions and there is no radiation of energy in extra dimensions (and maybe other auxiliary restrictive conditions)", and so forth. But all these theories are less and less attractive, although equally adequate from an empirical point of view.

For any scientific theory *T* no matter how empirically successful, there are countless alternative theories equally successful. Some are trivial, some are *ad-hoc*, and others are simply aberrant. How do we choose the right theory among countless rivals? Planck, Poincaré and more recently Popper all proposed in addition to empirical adequacy the

[76] Quine was one of the first to offer a description of theoretical virtues precisely in the context of underdetermination of theories and he paid special attention to simplicity (Quine 1963; Quine 1975).
[77] There are several ways of generating rivals for any theory. See (Maxwell 1998 51-53).

criterion of non-empirical considerations such as “elegance”, “simplicity”, “unity”, etc.[78] Are theoretical virtues, especially the virtue of unification, related to explanation, prediction and empirical success?[79] Moreover, if unification is a theoretical virtue, then an argument similar to the aforementioned one (mainly Quine’s) applies equally well to it.

B. Approach to unification by its instances and its practice. The second approach, closer in spirit to my proposal, pays special attention to the concrete instantiations of unification instead of approaching it at a general level. The “specific approach” is centered on exhaustive analyses of instances of unification and has been popular, particularly in the last decade. Some philosophers of science have discussed the most notable cases of scientific unification, focusing on its various virtues and drawbacks *in concreto*. Unification gets its license from facts and contexts, as well as from the beliefs of the scientists who managed to obtain it. It is then natural when philosophers and historians of physics examine concrete examples of unification, they find different concepts of unification at work. Depending on the case, such a specific approach to unification can successfully address the aforementioned problems of unification.

I prefer this approach to unification sensitive to content and specific, concrete instances of unification. I reveal the advantages of this type of unification and I argue for possible generalizations to other instances of unification. Instead of armchair speculations or absolute skepticism, the best method is to analyze cases of unification, to compare different definitions and accounts of unification, to accept and to praise and recognize cases of

[78] A more updated list includes at least the following theoretical virtues such as simplicity, unification, explanation, beauty, elegance, harmony, non-*adhocness*, coherency, invariancy, symmetry, organicity, perfection. See (Maxwell 2004).

[79] B. van Fraassen provided an argument to rule out any other virtues of explanation besides some pragmatic virtues: usefulness, simplicity, empirical adequacy (Van Fraassen 1977, 143-150).

"dis-unifications" or "dis-unity" in general. Morrison's book and subsequent discussions and reactions illustrate this approach at its best. I am sympathetic with her *a posteriori* way to address some issues such as: how unification is produced, unification's metaphysical implications and the role of unification in theory confirmation and theory choice. By acknowledging the complex nature of theory unification, she tried to unveil only *some* aspects of unification, maybe the most controversial ones (Morrison 2000 1sqq, 233). Although at the end I draw some general conclusions, my main aim is to start from facts and from the concrete practice of science and to see how unification was built up. But I am aware that leaning too much toward the "practice" can be misleading, too. I do not discuss the "psychology of unification" or the mental content of scientists who created unified theories. I prefer to avoid speculations on what Kaluza, Klein or others would have believed about unification.

From the analysis of the mechanism of unification "at work" a limited number of general claims can be drawn. I do not provide a general definition of unification and I do not claim to show that unification is *per se* necessarily linked to the progress of science.

In short, I belong to the enthusiast camp, but my approach is deflationary and similar to Morrison's. Unlike Kitcher, I do not approach unification at a general level, but I do not endorse Morrison's skepticism about unification.

6.2. The place of my case study within existing approaches

It is important to place Kaluza-Klein within the landscape of unification. The power of unification is not uniform across sciences. It is remarkable that in philosophical environments there are few studies of unification where it matters most—in post-relativistic theoretical physics. The absence of case studies in post-relativistic theories is detrimental to philosophy: if philosophy of science is to have relevance to science, its discussions should bear on the majority of work relevant to the topic. Lots of papers in philosophy of science have studied in detail unification in Newtonian gravity, Mendel and Darwin, Einstein and Maxwell. However, in many cases where unification is really taken as a special virtue—that is, in the attempts to unify the fundamental forces starting with Einstein—few philosophers have investigated it. Philosophy of science discussions of unification hence run the risk of having little to do with how unification is used in science where it is most used. I do not want to be too abrupt here: there are very comprehensive historical studies of the unified field theories and some of them exhibit deep philosophical insights, but none qualifies as a philosophical appraisal of the Kaluza-Klein theory.[80]

The known approaches have several aspects not discussed up to now or at least in work I am aware of. Although an inventory of the physics literature on Kaluza-Klein can take years to understand and to describe, I will show in Part III that the physics literature is deficient from a philosophical and from a historical point of view at the verge of being not only inaccurate, but misleading altogether. I hope, through a specific case study, I fill this gap by focusing on the concrete case study of Kaluza and Klein.

I go back to one place where modern unification started, the Kaluza and Klein attempted unification of gravity and electromagnetism within a formalism in which the spacetime manifold has five-dimensions. A study of unification in this area can thus reveal new insights into the way unification is designed, used and how it operates. These bring to

[80] For example, there are historical expositions of the unified field theories in (Pasini 1988; O'Raifeartaigh and Straumann 2000; Vizgin 1994; Wunsch and Goenner 2005; Duff 1995; Dongen 2002; O'Raifeartaigh 1997; Aitchison 1991); none of them discusses unification associated with the existing philosophical literature.

light new philosophical lessons. I argue that within the known Unified Field Theories, Kaluza-Klein makes the virtue of unification more plausible by analyzing the construction process of a physical theory that has the specific virtue of “theory unification”. It constitutes one of the most remarkable attempts of unification in physics, even if it is not the first and obviously will not be the last: subsequent programs include String Theory, Electroweak unification, various programs in Quantum Gravity, etc.

6.3. The relevance of Kaluza-Klein unification

I think of unification as creating a new theory T out of two existing theories T_1 and T_2. Arguably, the new theory has to be *different* enough from T_1 and T_2 and to bring about something new in the way we understand the world. In respect to other theoretical virtues, T needs to be more general and more explanatory than T_1 and T_2 taken together—or than their logical conjunction. This also implies that we agreed upon a method of counting phenomena and extimating the explanatory store of a theory—both being sometimes a complicated business. In my case, T_1 is General Relativity (**GR**) and T_2 is electromagnetism (**EM**) in its Lorentz invariant formulation. They were both empirical and epistemic successful theories, able to explain and predict different classes of phenomena that had been thought of as different “in nature”. Even if not explicit, in its various incarnations, Kaluza-Klein unification combines powerful ideas from geometry, group theory and quantum mechanics in order to provide the best theory in the sense of the epistemic virtues mentioned above. The hypothesis that spacetime has extra dimensions is supposed to dramatically increase the unificatory power of these two physical theories.

The reasons to look for unification in Kaluza-Klein theory are multiple. Unification is a special case of scientific progress and definitively a successful unification would improve our knowledge and understanding of the world, not always in the sense of providing new predictions or new applications, but providing better knowledge of the world. The theory was not motivated by the necessity to accommodate new, recalcitrant phenomena or to provide new predictions. Its main purpose was to provide a more coherent framework for T_1 and T_2. Moreover, the Kaluza-Klein model is prominent in string theory today; it provides a novel kind of unification based on the extension of space with extra spatial dimensions; it sheds light on the philosophical analysis of unification. I give here further reasons to analyze Kaluza-Klein theories:

Kaluza-Klein is a new type of unification. When we go back to Kaluza and Klein's theory, we shall learn much about unification itself. I argue that Klein's unification is not trivial and that it is not a reductive unification. In particular, we shall see how a new and distinctive type of unification emerges, one distinct from the case of Newton or Maxwell-Einstein. I will classify the Kaluza-Klein unification as close to the synthetic unification, but a new type of unification. Kaluza and especially Klein unified **EM** and **GR** in a specific way.

There is life after reduction for the unificatory programs in science. Because unification is related to theory production, I picture unification differently than reduction—although I do not claim that unification solves all the problems for reduction. I argue that Kaluza-Klein unification is non-reductive. Some suggest that unification is reduction at its best: I challenge this view. When unification succeeds, new and unexpected features of previous theories are revealed. Some unifications are based on mathematical identifications between the free parameters of previous theories: but in many cases unification goes beyond mere identifications and reductions: Klein's theory will illustrate this feature of unification. Even if one is tempted to think that Kaluza-Klein is a **GR** in five-dimensions, I will argue that after Klein's, the theory is not anymore the **GR** as formulated in **4-D**. I discussing Kaluza-Klein unification I stress its specific character, which is neither synthetic nor reductive, à la Morrison. Klein was able to introduce the wave-function in the process of unification and consequently was able to generate several novel and unexpected explanations.

Kaluza-Klein provides novel explanations. First, Kaluza-Klein unification brings about explanation. It provides *novel* explanations of various issues in theoretical physics: quantization of charge, features of the extra dimensions, internal symmetry of electromagnetism, etc. The main aim of Kaluza-Klein theory is to represent and explain two or more interaction forces under the same formalism/theory. It instantiates a special type of explanation that is arguably non-causal in nature but which can provide powerful explanations, including causal explanations. Although the explanation in Kaluza-Klein theory is not causal, there is room for causation in its later interpretations. For example, my case study can provide an answer to the age-old debate on whether geometry can explain and if so, how it explains and what it explains. In particular, it is a very controversial issue whether spacetime explains and in what sense it explains: under Mach's influence, Einstein, and lots of early physicists, thought that spacetime structure, an unobservable structure, cannot explain observational data and he called it “a factitious cause” (Nerlich 1994; Disalle 1995; Earman 1995). Kaluza-Klein theories shed light on a different question: is spacetime unificatory? Morrison, for example, won't agree that spacetime is the machinery of unification because seemingly there is no causation involved. As we saw in the previous sections, unificationists think that causal order follows explanatory order so in some sense they don't deny the existence of causal explanations; rather they just think they are not as fundamental.

Kaluza and Klein theories illustrate the stages of unification and of scientific progress. Kaluza-Klein is really two theories that are very different with respect to unification: the Kaluza's formulation and in Klein's formulation. I'll pay special attention to the way in which unification improved through this process. I show that Kaluza-Klein better organizes our knowledge about the world by providing a simpler and more unified way of systemizing known phenomena. The result is a better internal organization of scientific knowledge based on fewer principles and assumptions and on stronger connections between disparate phenomena. In Kaluza's theory, we see that many of the fears of the dissenters come to life. However, in Klein's theory, we see many of the virtues of unification represented; for in this theory, contrary to the suggestions of some dis-unifiers, unification is responsible for increased explanatory potential. In particular, we'll see that this is so because Morrison implicitly restricts to a controversially and narrow sense of explanation. Klein's theory has some unexpected consequences: it provides some new predictions, and moreover new explanations which are not causal or do not provide the machinery Morrison is looking for.

This approach is important for another perspective, too. A general question, unanswered up to now, is:

> ***[13]*** *How do we separate cheap/spurious/trivial unifications from substantive unifications?*

Is there a way to decide whether this unification here is good and that one is cheap spurious, trivial, etc? Such an approach needs to be comparative in nature. My case study puts an emphasis on comparing Kaluza to Klein and showing why Klein is better than Kaluza. But unification does not end up with Klein and definitively other projects need to be approached similarly, by comparison. Some authors suggest that when unification is couched in terms of group theory and symmetry groups there is a question to [13] (Maxwell 1998; Maudlin 1996, 129-144). Such an idea can in fact work for some theories which unify physical forces, but does not characterize well other unifications. For the time being there is no universal answer to [13], but I suggest here a way to look at several stages of an unificatory theory and compare them in respect of explanatory power, problem solved, *ad-hocness*, etc. And here is another suggestion: in answering [13], we may need to weigh differently explanation, prediction, coherency, etc depending on the unifying theory at stake.

Kaluza-Klein as a foundation for String Theory. Although it is the notorious forerunner to many unificatory attempts in String Theory, the Kaluza-Klein theory is barely mentioned as a case of unification in the philosophical literature.[81] A philosophical analysis of unification in String Theory should originate in the discussion of Kaluza-Klein. While Kaluza-Klein theory is known to be *ad litteram* false, heirs to it survive, and certainly the pattern of explanation is continually used today in the vast areas of String Theory

[81] (Weingard 1991, nd; Aitchison 1991) are among the few who discussed the Kaluza-Klein unification. Weingard explains why Kaluza-Klein is a special case of unification. Other hints to Kaluza-Klein as a unification can be found in (O'Raifeartaigh and Straumann 2000; Dongen 2002), but there is no extensive philosophical analysis in the literature.

and black hole theory.[82] In many ways it can be viewed as the harbinger of Einstein's unified field program, grand unified theories, and even superstring theory. There is also a controversial issue whether String Theory in its future formulations will retain the Kaluza-Klein mechanism. Some authors think that the trouble with String Theory is precisely the assumption of extra-dimensional manifold. Some others could argue that Kaluza-Klein does not play a special role in the foundation of String Theory. The majority of physicists would take it as one of the grounds of String Theory, albeit not the only one. This debate is underestimated in the philosophy of science.[83] I believe Kaluza Klein cannot be eliminated from the foundations of String Theory.

Other theoretical virtues and Kaluza-Klein unification. Last, Kaluza and Klein illustrate the connection between unification and other theoretical virtues. Simplicity, completeness, beauty, internal consistency, etc. are some of its virtues directly related to its most acclaimed virtue, its power of unification. But neither simplicity, nor "integration", nor "reduction" is sufficient to describe my case study. Each of them are present in the Kaluza-Klein theory but taken separately are not sufficient. Unification is conditioned upon other theoretical virtues, but it is not reducible to them. As M. Strevens proposed, if one is fond of a definition for unification, one can find other theoretical virtues as parts of its definition (Strevens 2004, 154-176).

In a nutshell, here are some general claims about unification that are illustrated by the Kaluza-Klein theories:

[82] The Kaluza-Klein particle and the Kaluza-Klein Ansatz, two aspects reminiscent of the original approach, are widely used in String Theory. In fact, Kaluza-Klein generated in the last decade a specific theory called the "Kaluza-Klein black hole theory". There are also attempts to explain gravitation as an aspect of Kaluza-Klein particles living in extra dimensions (Randall 2005, 499).

[83] I am currently working on a paper in which I deal with this issue and I will summarize it in Part III.

[14] *Unification can provide novel explanations and/or novel predictions not present in or not intended by the original theories. Unification is a major component of scientific progress.*

[15] *Unification can provide explanation, even in the absence of causality (or even when causality does not play a major role in the unified theory).*[84]

[16] *In many cases unification involves elements from other theories. Because there is a novel element used in the mechanism of unification, unification has a corrective aspect because more theories are involved.*

[17] *Unification is an inter-theoretic relation that is not reductive in nature. Or weaker: even the perfect unifications do not need to be reductive. Reduction and unification are separate inter-theoretic relations.*

[18] *The advancement toward unification is a multi-stage process of transformation of two theories* T_1 *and* T_2 *into a new theory* T.

[19] *Unification cannot be reduced to simplicity, generalization, cohesion, similarity, or analogy, although they are present and play major role in unification or its different stages.*

Summary of Part I

In the first part I made room for the discussion of the Kaluza-Klein unification by arguing for scientific unification and for its potential value for scientific progress. I exposed some recurrent themes (listed here in the order of their “strength”), frequently pondered in the philosophical literature on unification:

[84] In the case of Kaluza-Klein theories, there is a subsequent interpretation in which causal concepts are involved.

Theme I The inference from unification to confirmation and further to scientific realism;

Theme II The inference from unification to causation;

Theme III The relation between unification and explanation (or their "coupling");

Theme IV The connection between unification and other theoretical virtues (simplicity, generalization, cohesion, beauty, etc.);

Theme V The role of unification in the reduction-emergence debate .

In Chapter 1 and Chapter 2 I describe two main positions toward unification: the enthusiasts and the dissenters, by discussing the major controversies surrounding unification in the last three decades. I also expose two strong inferences from unification to confirmation/realism and respectively to causation (Theme I and Theme II above). In Chapter 3 I discussed the most popular accounts of unification, i.e. Friedman's and Kitcher's approaches to explanation as unification and the relation unification has with causation in Chapter 4. Chapter 5 is dedicated to a comprehensive discussion of Morrison's refutation of the strong inferences and relations as postulated in Theme I, Theme II and especially Theme III, based on her exhaustive analysis of counterexamples. In Chapter 6, I showed the intricacies of general approaches to unification in which it is defined *per se* or in terms of other theoretical virtues (Theme IV). I expressed my skepticism against any such approach and I took Morrison's side and supported her deflationary account, although I disagree with other important aspects of her approach, mainly the rejection of Theme III. I entered the specifics of my case study and I emphasized the novelty of my approach by anticipating the main lines of my analysis: Kaluza-Klein unification is neither a reduction, nor a synthesis of two theories and it is not a case of scientific emergence (Theme V).

Because of space limitations, I discuss only the most recent and the most popular approaches to unification. Furthermore, because of the peculiarities of my case study, I ignore a host of topics such as the connection between unification and confirmation in the Bayesian context (recently advanced by W. Myrvold) or the relation between unification and necessity, in C. Glymour's rendering (Schupbach 2005; Lange 2004; Myrvold 2003). In the present work, Theme I is not central–although in Part III, I shortly discuss the very touchy question of empirical confirmation of Kaluza-Klein theories (and of a larger class of theories that depend on the existence of extra spatial dimensions) in the context of scientific realism. I deal hereby with Theme II and adopt a pluralistic view of explanation in which unification can play a pivotal role and can provide causal-like explanations. Once one is pluralistic with respect to the theory of explanation and once one admits explanations that are not causal, unification can occur even when causation is not the dominant route to explanation.

PART II. KALUZA AND KLEIN THEORIES

In order to understand Kaluza's and Klein's attempts to unify the electromagnetic and the gravitational interactions we have to cast our minds back to the so-called "relativity years" (1915-1925).[85] I discuss the development of the field theories as a unificatory program during that decade in Chapter 7, more specifically the electromagnetic theory in its Lorentz invariant formulation (**EM**) and the relation between the two main ingredients of the Kaluza-Klein theories, **EM** and respectively the general theory of relativity (**GR**) in section 7.1. It is beyond the scope of my dissertation to give a philosophical presentation of **EM** and **GR**, as they were extensively discussed in the philosophical literature. I am interested only in: (A) their respective unificatory power and (B) the similarities and relations between **EM**, **GR** (and **SR**) as a condition to their unification. Given the importance of geometrization programs, I explore in Chapter 9 its philosophical aspects. Chapter 10 is dedicated to a comprehensive discussion of T. Kaluza's 1921 paper and Chapter 11 to a discussion of the two papers by O. Klein (both published in 1926).

[85] (Ryckman 2005)

Chapter 7. Unified theories at the beginning of the 20th century

Kaluza's and Klein's theories came about in a fortunate period in which precursory cases of unification—spanning over two centuries—were successful enough to boost physicists' enthusiasm: Newton's unification, Maxwell's unification of electric and magnetic fields and his unification of optics and electromagnetism, the theory of special relativity unifying space and time, it unifying dynamics with electromagnetism, etc. In trying to provide a unified description of reality, "mature" scientific theories endeavored to provide explanations, predictions and descriptions that had belonged to previous theories. By encompassing formerly accepted explanations and predictions, the new theories become more general. This gain in generality compared to previous theories had characterized both **EM** and **GR**, at various stages of their development: the formalism used was more general and they described more phenomena than before. Physical theories aspired to generality and universality in the sense of describing more phenomena. This process itself is related to unification as I show in the case of Kaluza and Klein. I do not claim that a theory needs to be general or mature in order to unify. Also, I do not think that *any* mature or general theory unifies. I do not look here for necessary or sufficient conditions of unification. In the concrete case of spacetime theories at the beginning of the 20th century, both **EM** and **GR** were mature and had a certain degree of generality. Obviously enough, there were not strictly speaking *the* fundamental theory. Lorentz formulation of **EM** was accomplished during the first two decades of the 20th century. **GR** was also a well established theory in the 1920s. When philosophers and historians of physics examined **EM** or **GR**, they found different concepts of unification at work as the two theories described two different interactions. The main aim of both Kaluza and Klein theories was to represent and explain two interaction forces under the same formalism and increase the generality of our representation of spacetime. Generalization is *one* among the components of unification: one of my aims is to show how Kaluza-Klein unification goes beyond mere generalization. I further elucidate the role of unification by comparing various aspects of unification present in the case of **EM**, **SR**, **GR** and Kaluza-Klein theories.

7.1. The covariant *EM*: an unificatory program

Before analyzing the Kaluza-Klein theory, I discuss its two components, i.e. the theories that are unified by it: **EM** and **GR.** I do not discuss **EM** and **GR** *per se.* There are several excellent philosophical analyses of both which go beyond the mere history of these theories and analyze philosophical aspects of both theories such as: determinism, locality, explanatory power, consistency, etc. It is interesting to note that these two components of Kaluza-Klein theories have in themselves a unificatory nature that plays a specific role in the genesis of Kaluza-Klein theories. I analyze Kaluza-Klein as unificatory programs. Except Morrison's analysis of **SR** and **EM**, there are few places in the literature where these theories are discussed as unificatory programs.

According to Morrison, the **EM** unification had two initial episodes: (a) the unification of electricity and magnetism on the one hand and (b) the unification of **EM** with optics on the other.[86] Morrison claims that the latter is more or less a reductive unification, while the former is more of a synthetic unification than a reductive one:

[86] Maxwell's two theories as exposed in his papers "On Faraday's lines of forces" (1856), "On physical Lines of forces" (1861-2) and "A Dynamical Theory of Electromagnetic Field" (1865) were extensively discussed in the philosophical literature as paradigmatic cases of unification.

> If we consider the structure of Maxwell's field equations, we cannot assume that the electromagnetic field reduces electricity and magnetism to one force. Instead, the electric and magnetic fields retain their independence, and the theory simply shows the interrelationship of the two—where a varying electric field exists, there is also a varying magnetic field induced at right angles, and vice versa. The two together form the electromagnetic field. In that sense the theory unites the two kinds of forces by integrating them in a systematic or synthetic way […] Yet there can be no doubt that the theory also reduced optical phenomena to their electromagnetic foundation. But without any *substantial* explanation of how that took place the reduction offered little in the way of *true* understanding (Morrison 2000 107 my emphasis).

For Morrison, both unifications are in fact less useful because they lack explanatory power. I take this quote as illustrating Morrison's thesis [5] (p. 111). For the present purposes, (a) is more relevant than (b).[87] According to Morrison, (a) is a synthetic unification because the **EM** field does not reduce electric and magnetic field, but unifies them through synthesis. In short, before Maxwell's first unification, magnetism and electricity were considered different classes of phenomena obeying different equations and having different explanations. After Maxwell's discovery, on almost all accounts, a huge number of phenomena were unified under the same theory. Besides these new explanations, Maxwell showed that the electromagnetic wave propagated in spacetime with a finite speed and that it was not an action-at-distance force, as Hertz had thought it to be. And this is another important achievement of the theory. Moreover, the propagation is made possible only by a

[87] I do not discuss whether the unification of optics and electromagnetism is reductive or not, although this is an interesting question. More precisely, optics reduces to a specific part of **EM** theory, the **EM** wave theory. I endorse the idea that it is more reductive in nature than the former unification. Some authors take it as a paradigm of reduction, similar to the reduction of Classical mechanics to SR by taking the limit $\frac{v}{c} \to 0$ in the Lorentz transformation or the reduction of thermodynamics to statistical mechanics when N, the number of particles of the system is $N \to \infty$. In the case of the reduction of optics to **EM** wave theory, the limit involves the Airy integral. But Batterman thinks that even in this case the reduction is not complete and a third theory, "catastrophe optics" emerges because the concept of interfering ray sums of the **EM** wave theory fails to capture the geometrical optics concepts of focals and caustics. (Batterman 2002 ch 6, esp. 88-90). As Batterman does not speak directly about unification though, I leave this topic aside for the moment.

reciprocal action of the electric and magnetic fields which retain their relative independence. For the unificationist, Maxwell's theory shows that electric and magnetic fields are nothing more than aspects of the electromagnetic field. In Morrison's interpretation there is a synthesis of **E** and **B** more than a reduction and I agree with her on this point.

For Nicholas Maxwell, synthetic unification is problematic in the case of electric and magnetic fields because "it is not obvious in general what must be done to show that *two* entities, or *two* forces, are really nothing but two aspects of *one* entity or *one* force." (Maxwell 1998 126) I see the problem N. Maxwell raises here and I have to admit that it is not easy to see this synthetic unification at work if J.C. Maxwell's theory is formulated in the language of **PDE**. Unfortunately, the concept of *one* force that unifies the electric and magnetic force does not do the unificatory job as it is not Lorentz invariant. We do not have the same situation as in Newton's unification where one force, the gravitational force, represented both the terrestrial and celestial forces. The best way to unveil the **E** and **B** fields as "aspects" of one entity is by shifting to the less attractive and less intuitive form of tensors.[88]

In addition, there is a fly in the ointment. Besides its beauty and simplicity, a deeper trouble of Maxwell's theory was its lack of a Galilean-invariant form. Maxwell equations were also not Lorentz covariant. It seems that Maxwell's theory unifies electric and magnetic field and gives a precise mathematical form to the electromagnetic waves, (which were explained by neither magnetism nor electricity), but it is at odds with **CM**. Some may even feel that **EM** unification creates a dis-unity with classical mechanics in the sense that

[88] It is merely a question of taste to think that Maxwell's equations as **PDE** are more beautiful and intuitive then tensors. Many physicists, Feynman included think that Maxwell's equations are elegant or even beautiful.

electricity and magnetism taken separately had been ‘more’ compatible with classical mechanics than Maxwell’s theory was. There were profound inconsistencies between the concepts of classical dynamics and electromagnetism, and Hertz and Fitzgerald had known the difficulties in using dynamics to explain electromagnetic phenomena. The two groups of invariances: Lorentz invariance and Galilean invariance were definitively at odds and they brought about disunity in the heart of physics (Morrison 2000 163). Electromagnetic fields do not obey the classical mechanics devised to describe particle for a simple reason: waves and fields are not particles and *viceversa*. But they need to coexist in the same world; composed systems of electromagnetic “objects” and classical particles needed a description, too. This inconsistency outlived Maxwell: Lorentz imposes invariance upon **EM**, but classical mechanics was not Lorentz invariant.

Another issue remains: because of its uneasy relationship with **CM**, Maxwell’s formulation of **EM** retained a duality between magnetism and electricity that gave rise to asymmetries in the case of electromagnetic induction. On the other hand, Maxwell’s theory was the first theory which suppressed the duality between matter and fields. Even without sources: charges, currents, densities of charges etc., the theory has non-trivial solutions, i.e. the electromagnetic waves. The electromagnetic field lives in empty space without matter or charges. In classical mechanics, the duality between matter and field was present, because Newtonian theory talks about point particles, fluids, densities of mass, etc. and the empty space imbued with forces.[89]

[89] There is a way of interpreting Newtonian mechanics as a pure field theory by adding a “field of impenetrability” to all the other forces discussed in *De Gravitatione* and in the *Optiks*. This position was advocated in (Stein 1989; 1970, 384) and more recently in “Newton’s Metaphysics” in (Cohen and Smith 2002, 500). I critically discussed this interpretation in an unpublished paper (Muntean 2005).

Another unification was necessary. It was Einstein's achievement to show that the laws of electrodynamics were valid in all frames in which the laws of mechanics held. As it is generally admitted now (rejected only by small class of dissenters), Einstein was able to unify the dynamics of massive bodies with electromagnetism of charged particles and fields by creating a new theory that was internally consistent or at least more consistent than **CM** and **EM** were. The undesirable inconsistency with mechanics was still present in Lorentz such that the unification of electric and magnetic field was accomplished only after Einstein's 1905 papers. One can say that he removed the duality between electric and magnetic fields by providing a Lorentz invariant formulation of electromagnetism. The duality is better described in a later work of Einstein:

> Take, for example, the reciprocal action of a magnet and a conductor. The observable phenomenon here depends only on the relative motion of the conductor and the magnet, whereas the customary view draws a sharp distinction between the two cases in which either the one or the other of these bodies is in motion. For if the magnet is in motion and the conductor at rest, there arises in the neighborhood of the magnet an electric field with a certain definite energy, producing a current at the places where parts of the conductor are situated. But if the magnet is stationary and the conductor in motion, no electric field arises in the neighborhood of the magnet. In the conductor, however, we find an electromotive force, to which in itself there is no corresponding energy, but which gives rise—assuming equality of relative motion in the two cases discussed—to electric currents of the same path and intensity as those produced by the electric forces in the former case (Einstein 1923 37).[90]

Why was Maxwell's **EM** inconsistent? In their most elaborated form, Maxwell equations described how the electric field **E**, the magnetic field **B** and the **EM** waves (radio

[90] The same idea is present in his "On the Electrodynamics of Moving Bodies", 1905. According to J. Norton, this inconsistency between CM and EM led Einstein to the formulation of **SR** (Norton 1985, 203-246)

waves as well as light waves) behave in a vacuum or in the presence of charges and currents:

$$\begin{cases} \nabla \cdot (\varepsilon \mathbf{E}) = \rho \\ \nabla \cdot \mathbf{B} = 0 \\ \nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t} \\ \nabla \times \left(\dfrac{1}{\mu} \mathbf{B} \right) = \mathbf{J} + \dfrac{\partial (\varepsilon \mathbf{E})}{\partial t} \end{cases} \tag{7}$$

here ε is the electrical permittivity, μ is the magnetic permeability, $\mathbf{E} = (E^1, E^2, E^3)$ is the electric 3-vector, $\mathbf{B} = (B^1, B^2, B^3)$ is the magnetic 3-vector, **J** is the current density or the ordinary vector of currents and ρ is the charge density. By definition, $\rho = 0, \mathbf{J} = 0, \varepsilon = \varepsilon_0, \mu = \mu_0$ in vacuo.[91]

Something is missing in (7), which describes the dynamics of the **E** and **B** fields by terms such as: $\dfrac{\partial \mathbf{B}}{\partial t}, \dfrac{\partial \mathbf{E}}{\partial t}$. What one needs on top of the dynamics of **E** and **B** is (a) the dynamics of charges, conductors, magnets, etc. *and* (b) the dynamics of observers. Beseides not being Galilean invariant, Maxwell's theory violated the relativity principle formulated by Einstein. The concepts of "movement of charges" and "movements of magnets" had been known to Faraday and Ampère and they are paramount to Maxwell's theory, but they were not invariant to the change of system of reference. The incompatibility with dynamics lies in the way **EM** represents the motion of inertial systems of reference, i.e. the observers of electromagnetic phenomena, not those of sources, i.e. charges,

[91] Hereafter, I designate the spacetime directions by numerical indices instead of the Cartesian notation x, y, z and t. B^1 is the **B**'s component on the first axis. Time (t) is the zeroth component of a 4-vector. I follow notations and conventions used in (Carroll 2004).

conductors, and magnets. And here Maxwell's theory gave very unrealistic results, as it is not Galilean invariant. Among other issues, the spherical form of a front wave was not an invariant of the theory. Maxwell theory predicted that in different systems of reference the wave deformed from a sphere to an ellipsoid, hyperboloid, etc. depending on the ratio between a system's velocity and the speed of light. This was one of the worst predictions of the theory that contradicted the existing empirical data.

Given all these major incompatibilities between **CM** and **EM**, how should one expect unification? The conjunction of electromagnetism & relativistic dynamics is inconsistent, so a change is needed in **EM**. That change was writing **EM** in a covariant formulation.

There is an elegant solution to the inconsistency between the dynamics of fields and the dynamics of charges, conductors magnets and observers inspired by Minkowski's assumption. He postulated a description, geometrical in nature, of a 4-D manifold, "the Absolute World" and added the fourth coordinate to the three existing space-like coordinates. In **SR** we need always four numbers to represent *events* in spacetime, or in other words, the physical spacetime is coordinatizable by $\mathbb{R}^4$. In the oft quoted passage from his 1908 talk and 1909 paper, Minkowski claimed: "[…] space by itself, and time by itself, should be completely reduce to shadows, and only a kind of union of the two will preserve an independent reality" (Minkowski 1910 esp. part II). Beside the union between space and time, adding a new coordinate to the three spatial coordinates has in itself a unificatory power especially for the **EM** theory: it would help Lorentz, Minkowski and finally Einstein to realize the unity of electric and magnetic fields.[92]

[92] The ideas discussed in the section do not belong to one author. Minkowski, Abraham, Planck, Lorentz, Laue and Einstein himself contributed to the tensor formulation of **EM**. It seems that Minkowski was the first

A prototype of a 4-vector is the displacement of positions $\Delta\mathbf{r} = \left(c\Delta t, \Delta x, \Delta y, \Delta z\right)$.

Similarly, there is a 4-velocity: $\mathbf{u} = \dfrac{d\mathbf{r}}{d\tau} = \left(c\dfrac{dt}{d\tau}, \dfrac{dx}{d\tau}, \dfrac{dy}{d\tau}, \dfrac{dz}{d\tau}\right)$ (where τ is a parameter) and a 4-acceleration **a**, etc. Other mechanical quantities can be generalized to 4-vectors. Once this step is taken, the derivative in time and the derivative in space can be expressed in a unitary way.

In Minkowski 4-D space we can perform a calculus of vectors similar to the calculus of 3-vectors, but the 4-vectors have a wonderful property: they are invariant under Lorentz transformation. If one wants the laws of physics to be invariant in all inertial frames, then the 4-D formulation is the only one that passes the relativity principle as required by Einstein :

> From the totality of natural phenomena it is possible, by successively enhanced approximations, to derive more and more exactly a system of reference *x, y, z, t*, space and time, by means of which these phenomena then present themselves in agreement with definite laws. But when this is done, this system of reference is by no means unequivocally determined by the phenomena. *It is still possible to make any change in the system of reference that is in conformity with the transformations of the group G_c and leave the expression of the laws of nature unaltered* (Lorentz and others 1952 79)**.**

How are 4-vectors generated from three vectors? Usually, a correspondence with 3-D mechanics is the strong rule of thumb: pick the spatial component of the 4-D vector as being identical with the 3-D vector. But one has some freedom in choosing the fourth element. To a 3-vector we add an element which is inferred from laws of conservation or

---

who wrote the tensor form of **EM** in a conference delivered in December 1907, but did not publish it till 1910.

by sneaking a peak at the **EM** theory in its Minkowski form.[93] In some situations, the three vectors can be misleading. For example, in choosing an expression of the 4-force F, one can take inspiration from two relations: either from $\boldsymbol{F}=m_0\boldsymbol{a}$ or from $\boldsymbol{F}=(d/d\tau)\boldsymbol{p}$**.** The latter is preferred for reasons unrelated to the classical mechanics, rather to the wave equation wave or to Lorentz transformations. In fact, relativity strived to replace $F=ma$ with a formulation of a *continuum* mechanics in terms of stress-energy momentum tensors. For Einstein, this was one of the most important advances in relativity before **GR**.

A convenient way to write the 4-velocity is:

$$\mathbf{U} = \gamma(u)\left(\mathbf{u}, c\right) \tag{8}$$

so a way of writing the 4-force is:

$$\mathbf{F} = \gamma(u)\left(\mathbf{f}, \frac{1}{c}\frac{dE}{dt}\right) \tag{9}$$

where the 3-D force **f** is defined in a relativistic fashion as: $\mathbf{f} = \dfrac{d(m\mathbf{u})}{dt}$ and *not* as $\mathbf{f}=m\boldsymbol{a}$.

It is also true that 4-vectors are not enough to express all laws of physics: **EM** needs 4-tensors. The choice of the 4-tensors is not obvious because the number of elements to be added to the three existing components and the way to match 3-tensors into 4-tensors is not unique. By the way of an analogy with the 3-D **EM**, one wants to associate a (2,0) tensor to the fields and a (1,0) to the potentials. On top of this, another problemis that the 4-potential is not uniquely determined by the 4-tensor of the electromagnetic field $F_{\mu\nu}$. This will later

---

[93] I briefly refer to Minkowski's work here. His lecture *Space and Time* delivered in September 1908 had a huge impact and it is well described in the philosophical literature on **SR;** see translation in (Lorentz and others 1952). Recently, in a paper presented at the *Third International Conference on the Nature and Ontology of Spacetime* (Montreal, June 2008) M. Janssen and R. Rynasiewicz it is argued that Minkowski's construction is not, as usually believed, just "simply a matter of inserting a few factors of gamma into F=ma to produce a Lorentz-invariant version of Newton's second law". See: http://www.spacetimesociety.org/conferences/2008/cprogram.html

raise a host of philosophical interpretations associated with it, the so-called "gauge theories".

The first formulation of the **EM** in tensor form implied some patchy guesswork on Lorentz' and Minkowski's behalf. The mathematics to be used was not completely developed and effective at that time.[94] In order to emphasize this guesswork, it is enough to see that the tensor $F_{\mu\nu}$ was defined on the premise that the 4-force should be "linearly" dependent on the 4-velocity U^{μ} based on the 3-D equation derived from the standard Maxwell equation:

$$\mathbf{f} = q\left(\mathbf{E} + \frac{\mathbf{U} \times \mathbf{B}}{c}\right) \tag{10}$$

The quotient of this relation should be the magnetic field. The guessed equation for the force was:[95]

$$f_{\mu} = \frac{q}{c} F_{\mu\nu} U^{\nu} \tag{11}$$

where one can see why $F_{\mu\nu}$ plays the role of electromagnetic field in 4-D. The potential vector A_{μ} was inferred from the premise that there is a quantity whose curl is $F_{\mu\nu}$. There is no "curl" in tensor calculus, but the closest operation is the antisymmetrized derivative:[96]

$$(\vec{\nabla} \times \vec{F}) = \hat{\mathbf{e}}_k \epsilon_{k\ell m} \partial_{\ell} F_m \tag{12}$$

The fact that the magnetic and electric phenomena are dependent on the system of reference is directly related to the inconsistency between relativistic dynamics and electromagnetism. It is a commonplace to claim that both were eliminated in Einstein's **SR**. In

---

[94] *Pace* Wigner (Wigner 1967, 280)

[95] Here the Greek index indicates the component of the force. It has values from 0 to 3.

[96] Here I did not faithfully and systematically follow the work of Minkowski or Lorentz. A more elaborate discussion can be found in (Rindler 2006 ch. 7).

the wake of the relativity years, in "Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern" in (Speiser and Weyl 1911 386), Minkowski showed that the components of electric and magnetic fields transform into each other under the group of Lorentz transformations. Rather than a collection of two vectors **E** and **B** with six seemingly independent components, the electromagnetic field in **SR** is described by the "field strength tensor" $F_{\mu\nu}$: the (0,2) differential form having the following components:

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix} \tag{13}$$

It is related to the 4-vector potential $A_\mu = \begin{pmatrix} V & A^1 & A^2 & A^3 \end{pmatrix}$, where V is the scalar potential, and to its first derivatives:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{14}$$

All Maxwell equations (7) can be derived from covariant equations:

$$\begin{cases} \text{inhomogeneous: } \nabla_\nu F^{\mu\nu} = -4\pi J^\mu \\ \text{homogeneous: } \partial_{[\mu} F_{\nu\rho]} = 0 \end{cases} \tag{15}$$

where the 4-current is:

$$J^\mu = (c\rho, \vec{\mathbf{J}}) \tag{16}$$

The homogeneous equation is simplified to:

$$\partial_\mu F_{\nu\kappa} + \partial_\nu F_{\kappa\mu} + \partial_\kappa F_{\mu\nu} = 0 \tag{17}$$

Tensors in the covariant **EM** in such a convoluted expression are needed, instead of the equation of vectors, because unlike kinematical vectors and scalars such as velocity, momentum, energy, etc., electrical and magnetic vectors do not transform under Lorentz transformations. If V^{μ} is a 4-vector (position, momentum, velocity, acceleration) in a system of reference S_1, then its components $V^{\mu'}$ in another system of reference S_2 are related such that:

$$V^{\mu'} = \Lambda_{\nu}^{\mu'} V^{\nu} \tag{18}$$

where the matrix Λ encodes the transformation of the 4-coordinates and 4-velocities from what S_1 measures to what S_2 measures and *vice versa*. For a "boost" along the first direction (i.e., the x axis) with a velocity defined as v=ctanhφ, Λ is given by:[97]

$$\Lambda_{\nu}^{\mu'} = \begin{pmatrix} \cosh\varphi & -\sinh\varphi & 0 & 0 \\ -\sinh\varphi & \cosh\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{19}$$

The central problem is that **E** and **B** simply do not obey (18) or anything remotely similar to it. In the flat spacetime, the kinematics of point particles does not need tensor calculus whatsoever, whereas even in flat spacetime, the dynamics of **E** and **B** cannot be written without using tensors.

Is the tensor a convenient notation? Tensors are convenient notations that encode components like complex numbers, vectors or matrices. My purpose here is to show that encoding two or more vectors or components in the same mathematical structure is not unification. It is easy to provide here a counter-example, the complex **EM** field, known long before Lorentz. Because **E** and **B** fields display a so-called "duality", we can represent them by a complex field:

$$\mathfrak{E} = \mathbf{E} + i\mathbf{B} \tag{20}$$

such that Maxwell's equations are written simply as:

[97] I use the convention c=1 in this equation and it will be implicit is some of the subsequent equations.

$$\begin{cases} \nabla \cdot \mathfrak{E} = 0 \\ \nabla \times \mathfrak{E} = i \dfrac{\partial \mathfrak{E}}{\partial t} \end{cases} \tag{21}$$

These equations encode in a very simple manner all that Maxwell's equation represent. They help in finding a solution to the wave equation, etc. It is also simpler and maybe more economic than (13) and definitively more elegant, simpler and more economic than (15). But is it unification in a non-trivial sense? I argue that unification goes beyond mere simplicity and beyond better codification and that we have reasons to prefer (15) to (21). The dissenter could say: (21) is unification, no more or no less than the unification achieved by the tensorial calculus in (15). My response is that there are several things missing from the "unification" depicted in (21) compared to that achieved in (15).

One can take tensors as structures that represent electromagnetism and dynamics in the 4-D world more seriously than any "pasting" structure such as (20). We can unify two completely separate theories by inventing a tensor product for child psychology and QCD for example:

$$\begin{pmatrix} \text{the dynamics of child psychology} & \text{(presumably) } 0 \\ \text{(presumably) } 0 & \text{QCD} \end{pmatrix} \tag{21}$$

It is clear that with a convenient mathematical notation we can paste anything anywhere and Feynman pitiful attitude toward boastful unification is perfectly justified.[98] But as I suggested in the Part I, in unifying two theories there is more than pasting and conveniently locating components. The tensor is well behaved under transformations of coordinates between inertial systems—which is a mechanical feature, whereas (20) is not.

[98] See (Maudlin 1996, 129-144; Feynman, Leighton, and Sands 1989) and Section 12.6 in the present dissertation.

If we adopt a weak holism here, tensor calculus fits better within the network of other theories than a trickery such as (20). A simple analogy with spatial objects can be useful here. The projection of a cube on a plain can be a square, a rhomb or a more complicated polygon, depending on the perspective. Insisting that one representation of a cube is more fundamental is hapless. Some give more information about the cube, some less. This "difference" depends on the perspective and is pure representational. But there is something that unifies all these planar representations is the cube itself with its invariants to several representations. I think that a complex vector like (20) is nothing more than a combination of planar projections whereas the tensor (14) is closer to the cube in the previous analogy.

Why do we need tensors in order to unify **E** and **B** fields? It is not that they are simpler or more elegant than other representations, but because tensors are invariant in a desirable way. In fact, the same argument from the unity with the dynamics works here, too. Tensors fare much better than 4-vectors: there was no way to formulate **EM** in a covariant form with vectors only. Indeed, in two different inertial systems the components of $F_{\mu\nu}$ are well transformed under a Lorentz transformation:[99]

$$F_{\mu'\nu'} = \Lambda^{\mu}_{\mu'}\Lambda^{\nu}_{\nu'}F_{\mu\nu} \tag{22}$$

This is then the sought for covariant formulation of **EM**, analogous to (18). The moral to be drawn from this is that what is measured as "electric" in S_1 can become what is measured as "magnetic" in S_2 and vice versa. One can see that the reductive unification criticized by is plausible on this interpretation.

[99] $F_{\mu'\nu'}$ are the components of the field strength tensor in S_2.

The difference between electric and magnetic components of the vectors is not invariant anymore because it depends on the system of reference. In the first decade of the 20th century, this interpretation of electromagnetism became the norm. According to the invariance principle proposed by Weyl (Weyl 1952 132) and adopted by Einstein in the case of ether, this difference between the "measured" **E** and **B** is not absolute anymore, as it depends on choice of coordinates (Kosso 2003 414). Other theoretical entities have had similar fate: the "ether" was downgraded by **SR** from the status of an absolute object to the status of a relative object; acceleration stayed absolute in **SR** but became relative in **GR**; in Riemannian **GR**, only $g_{\mu\nu}$ was absolute. The ontological division between electric and magnetic realities is overtaken by the formalism because the difference between electric and magnetic realities is frame-dependent and can be considered an artifact of our description. The status of $g_{\mu\nu}$ in **GR** is analogous: it is an invariant object, although its components are not.[100] However, according to Morrison, this doesn't mean that electric and magnetic fields are reduced by elimination one to the other. For a given system of reference, electric reality can be neatly separated from the magnetic reality: in a different system of reference, the measured values of the **E** and **B** fields can be flipped. It is clear that the electromagnetic field is better represented by a "2-form", i.e. the $F_{\mu\nu}$ tensor than the disconnected and "pasted pair". These are some computational reasons to prefer tensors to the "pasted pair". But there is a subtler reason to take tensors more seriously.

Remember that spacetime and reality does not come equipped with a system of coordinates. What is given is matter, fields, and currents, etc. in spacetime. As I see it,

[100] This follows from Einstein's "equivalence principle" in which the difference between inertial and gravitational mass is not fundamental because it depends on the system of reference. Thanks to C. Wüthrich for emphasizing this aspect.

tensor calculus is a step away from the coordinate systems because although their components are coordinate dependent, as a *whole* they have several invariances to the choice of coordinate systems. Tensor calculus, despite its complications, is a step towards reality and not towards mathematical fictions. The pair such as (E, B) or the complex vector (20), although seemingly more elegant than tensors, do not display invariance and are coordinate dependent in a "bad way". And if invariance is a criterion for reality, then tensors are more real than vectors.[101] The missing link is between invariance and explanation. It is easy to see that invariance and conservation are in fact explanatory in **EM**. This argument is not discussed by Morrison, although she suggests the other direction of argument according to which the tensor calculus pushes us away from explanation, despite the host of arguments based on symmetry and invariance purporting to show that tensor calculus is indispensable in describing reality.

There is an independent reason to entertain a difference between **E** and **B** within **EM** that has nothing to do with tensors or explanatory structures: the different nature of the sources of electric and magnetic fields. Morrison does not discuss this point which is at this stage an empirical question. The empirical aspect of this asymmetry is the existence of electric monopoles (charges) and the inexistence of magnetic monopoles. There are magnetic dipoles and electric monopoles and dipoles. In this respect, the real world makes a difference between the sources of electric field and the sources of magnetic field. We do not observe magnetic monopoles (yet?). The experimental evidence for their existence is

[101] Usually Einstein is credited for stating invariance as a criterion of reality. I am more interested in invariance as a feature of unification.

controversial, at best.[102] In the absence of magnetic monopole, the way we detect **E** and **B** fields with probes is different. We cannot detect the **B** field with a particle at rest: a magnetic monopole would permit this detection. One can see here a more serious trouble: it may be the case that they are generated in two different ways by different dynamics. And in this respect, from a dynamical point of view, Maxwell's **EM** is still a theory that retains a special difference between electric and magnetic dynamics.[103]

7.2. EM as a synthetic unification (N. Maxwell and M. Morrison)

These reasons suggest that we deal with a different type of unification than the reductive one. Morrison and N. Maxwell both suggested that synthetic unification is at work in the case of **EM** theory. According to N. Maxwell, the new theory *T* is a synthetic unification if it meets one or both of the following conditions (Maxwell 1998 130):

> *(a)* T must show how "the distinct entities or forces $E_1 \ldots E_N$, interact with *one* another in a symmetric way, so that the existence *of* any *one of* E_1 to E_N implies the coexistence of all the others." and
> *(b)* T must show that "the manner in which the *unified* entity or force, *E*, splits up into the distinct entities or forces, $E_1 \ldots E_N$, depends on nothing more than the adoption of an arbitrary convention from a range of equivalent possibilities—such as in the case of the electromagnetic field, the adoption of one reference frame from infinitely many other, equally good reference frames in uniform relative motion with respect to each other.

The **PDE** formulation of **EM** which is non-covariant meets none of Maxwell's criteria and the theory is blatantly inconsistent with the motion of material particles and

---

[102] The detection of magnetic monopole, announced in 1975 but not confirmed since then, is a major subject of controversy between String Theory and the Grand Unified Theories. I am not sure whether this shows that the difference between **E** and **B** is as powerful today as it was at the beginning of the 20th century. P. M. Dirac's proof of 1931 that electrical charge and magnetic charge are quantized could have broken this difference.

[103] It is a very interesting question whether **QED** solves or bring something new to this problem.

inertial systems of reference. Hence one needs to abandon the original formulation in favor of its covariant formulation.

A final point on the **EM** unification: Morrison and I agree on its unificatory power. Arguably it is the exemplar unification, but what makes it non-trivial and non-spurious? I provided here a trivial unification: the complex vector $\mathfrak{E}$ "unifies" **E** and **B**, but in a trivial way. Think of another extreme example: a theory saying that "**E** field and **B** field are both subjected to God's will" is also unificatory. Both are trivial for different reasons, whereas Lorentz unification is not. I try to define the **EM** type of unification in order to set the standard of the perfect unification. In order to define it, we need the concept of Lorentz transformation (call it L-transformation). Here is a definition:

Def 8 Unification as a L-transformation*: Let Q_1 be a quantity described in a theory T_1 by a mathematical structure $\mathfrak{S}_1 = A_1^{(1)}...A_1^{(N)}$, and Q_2 another quantity belonging to a theory T_2, described by the structure $\mathfrak{S}_2 = A_2^{(1)}...A_2^{(N)}$. If $\mathfrak{S}_1$ and $\mathfrak{S}_2$ can be integrated in a tensor F (having all the properties of tensors in Minkowski space-time), such that under some Lorentz transformations can be changed one into each other by a function Φ:*

$$\mathfrak{S}_2 = \Phi(\mathfrak{S}_1) \tag{23}$$

then quantities Q_1 and Q_2 are unified *under the L-transformation.*

Given (23), Φ is not arbitrary and admits an inverse, so likewise Q_2 can be transformed back into Q_1.

In **SR**, the following pairs are unified in the sense of **Def 8** within 4-vectors and tensors:

- time is unified with space as represented together in X^μ;
- energy is unified with momentum as represented together in p^μ;
- characteristics of matter (pressure and density) are unified as represented together in the energy momentum tensor $T^{\mu\nu}$;

SR had the great merit of using electromagnetism in order to endow Lorentz transformations with a realistic interpretation. Although momentum and energy refer to different realities that can be reduced neither as in reductive cases discussed below, nor in the sense of L-transformation, unification is possible in a weaker sense than **Def 8**. We do not expect that all unification work as such. But they retain their individuality from a metaphysical point of view although for reasons which are not necessarily related to Morrison's. They are represented unitarily within one and the same mathematical structure.

L-unification in the sense of **Def 8** is not reductive, but synthetic. R. Weingard held that this kind of unification is the strongest and it is *not* present in the Kaluza-Klein theory (Weingard 1991, nd). This does not mean that Kaluza and Klein do not unify, but that they unify in a different way than the L-transformation. I can say that the Kaluza-Klein unification has a prepoderent synthetic character in which both gravity and electromagnetism cannot be reduced one to the other, but they are aspects of a different kind of interaction in five-dimensions.

There is a sense in which Unified Field Theories try to extend such a definition by using the concept of "hyperfield". For Lichnerowicz, a theory *T* unifies "in a broad sense" two fields in one field if it attributes symmetrical roles to the two fields; in **GR** for example, the two fields should result from the same same geometry (Lichnerowicz 1955, 298).

When unified in the "broad sense", the theory retains the reality of the two fields. On the other hand, a theory is unified in "a strict sense" if the exact equations govern a *non-decomposable* hyperfield, and they can only *approximately* be decomposed into two field equations when one of the fields dominates the other. So in Lichnerowicz' definition, the strict unification is ascertained when the previous, distinct fields are approximantions of the "hyperfield".

The same question can be asked in the case of **EM**. Were **E** and **B** ontologically demoted once $F_{\mu\nu}$ was discovered? There are two possible answers here: the straightforward one is that F is more real than **E** or **B** which are not anymore *objectively real*. We see **E** or **B** when we pick certain reference frame. **E** and **B** are eliminated from the primary ontology by being replaced by a frame-independent entity F. Maudlin endorses this answer when he paraphrases Minkowski's remark: "the electric field by itself and the magnetic field by itself are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality" (Maudlin 1996 133). The other answer, suggested by Morrison, is that the **E** and **B** are not absolutely eliminated. I avoid such a conclusion because it goes well beyond the **EM** unification itself into the domain of gauge invariances and its ontological commitments. It worth noting that from the previous arguments based on unification one could infer that the unificatory structure F is more real than the unified realities Q_1 and Q_2. If $Q_{1,2}$ fades away or not is a complicated matter that involves lots of metaphysical assumptions—similar to the case of time and space involved by Maudlin. I will ask a similar question in the context of Kaluza and Klein theory.

Chapter 8. How to mesh together *EM* and *GR*?

GR is a more powerful theory than **SR** in the sense that it deals with more general structures of spacetime. Can we talk about the unity of **GR** in the sense discussed in Chapter 1 (p. 10sqq.)? According to the defi**ni**tions I **u**se, what is precisely unified in **GR**? Does **GR** constitute unification of previous theories, and if so, **w**hich theories? **F**inally, in what degree is **GR** connected to **EM**? Do they independently coexist or, on the contrary, one would expe**ct** them to **b**e unified? What is the place of **EM** within **GR**, if any?

There are indeed some aspects of **GR** that conjure up unity: the dynamics of the spacetime metric and the general covariance of the field equations. The former states that the metric of spacetime is dynamical as in any other fields: it depends on the distribution of masses as in the case of other fields. The latter is a more technical aspect to be discussed later. But this is not the end of the story in respect to how unified **GR** is. My task here is to discuss **GR** as an unifying theory in the meaning discussed in Part I.[104]

According to Einstein's "equivalence principle", **GR** unifies inertia and gravitation in "a logical unit" (Einstein 1929 127), similar to the way **SR** had unified mechanics with electrodynamics or energy with matter.[105] Despite this popular **GR** mantra, the unificatory contribution of **GR** is not that clear. Classical mechanics had treated the two masses as equal, but still as two separate entities. By its principle of equivalence, **GR** assumed that they are identical and there is no experiment that could have made the distinction between

[104] In the period 1915-1921, some members of the relativity community (Einstein included) were looking at **GR** as a unifying theory. I discuss **GR** only them as long as it serves the purpose of my case study. Historical analyses of **GR** as a unificatory program can be found in (Ryckman 2005; van Dongen 2002).

[105] I refer here also to a text published in 1920 under the title "Äther und Relativitätstheorie" which is based on the address delivered in Leiden on May 5, 1920. See the translation in (Renn 2007 613-619). The address is available online: see (Einstein 1920).

an accelerated system, i.e. the inertial mass of a probe particle, and a system subjected to a uniform gravitational potential, i.e. its gravitational mass. Second, despite this major step forward, there are other dualities that Einstein wanted out of his theory. He wanted to remove, or at least reduce, the gap between the so-called “matter-ether duality” or the “spacetime-matter opposition”:

> Since according to our present conceptions the elementary particles of matter are also, in their essence, nothing else than condensations of the electromagnetic field, our present view of the universe presents two realities which are completely separated from each other conceptually, although connected causally, namely, gravitational ether and electromagnetic field, or—as they might also be called—space and matter. Einstein (1920) in (Renn 2007 619).

A few lines down, the unificatory ideal in Einstein’s is described as: “it […] would be a great advance if we could succeed in comprehending the gravitational field and the electromagnetic field together as one unified conformation. […] The contrast between ether and matter would fade away, and, through the general theory of relativity, the whole of physics would become a complete system of thought, like geometry, kinematics, and the theory of gravitation.”

In fact, **GR** itself did not accomplish that unification. Maybe the renowned dream of Einstein, to unify matter and fields, is an impossible ideal even today. Third, there was the unification of **EM** and **GR**. What if we take the EM field as a field *per se*, without bearing a relation to any form of matter and try to unify it with the gravitational field? Einstein was not satisfied with the way **GR** treated **EM** field either (Pais 1982). This unification is associated more with names like Weyl, Hilbert and Eddington than with Einstein—despite his obsession with unified field theories after the 1930s. **EM** is present in **GR** because there is a specific part of the energy-stress tensor that corresponds to the energy carried by **EM** fields and charges.

There were slim hopes that gravity and electromagnetism could be unified in the sense given by "invariance" in Def 8. There is a conceptual problem with extending it to gravity, because **GR** goes far beyond Lorentz transformations. In fact, **GR** does have Lorentz transformations only locally and only in a certain approximation. One needs to drop the ladder of Def 8 and look for something more substantial. Indeed, in 1913, Einstein realized that gravity, unlike **EM**, cries for a description in which the dependence of coordinates on the metric is not anymore linear (Einstein 1913, 487-500). Even the concept of inertial system of reference is ill-formed in **GR**. Either one drops the Lorentz transformation condition in **Def 8** or one looks for another, weaker sense of unification. Definition **Def 8** simply does not apply to **GR-EM** unification either. We do not want to convert gravitational field into electric or magnetic fields in different systems of reference. There is no empirical evidence that gravity becomes electromagnetism or vice versa by merely changing the reference frame. Is it possible to seek another type of unification, weaker than Def 8? In many respects, Kaluza and Klein came with a new, different type that is not related to the Lorentz unification. In order to do this, the aspiration to a reductionist unity in the sense of the simplistic "electromagnetic program" and the hopes for a "L-unification" should be altogether dropped. In the same vein, the unification sought by Kaluza and Klein is weaker than what one finds in Einstein's **SR**. So there is a partial answer to the pressing question [13] (p. 136). The question is whether it works and whether it provides explanations, albeit being weaker.

At the core of the **GR** one can find Einstein's "Equivalence Principle" stating that in small regions of spacetime the laws of physics reduce to those of special relativity and there is no way to detect the gravitational field itself by means of local experiments.[106] One of the consequences is that the interaction of matter fields to curvature is minimal and there are no direct couplings to the Riemannian tensor or contractions. The other consequence of the equivalence principle is that there is a general covariant form of all laws of physics, i.e. the Principle of Covariance. Very roughly, the principle of covariance says that a theory *T* can be reformulated in curved spacetime by:[107]

a) replacing the ordinary derivative with covariant derivatives:

$$\partial_{\mu} \xrightarrow{GR} \nabla_{\mu} \tag{24}$$

b) replacing the Minkowski metric η with the metric g:

$$\eta_{\mu\nu} \xrightarrow{GR} g_{\mu\nu} \tag{25}$$

For example, **EM** is rewritten in curved spacetime by replacing the ordinary derivative with the covariant derivative in Maxwell equations:

$$\begin{aligned} \nabla^{\mu}\mathbf{F}_{\mu\nu} &= -4\pi\mathbf{J}_{\nu} \\ \nabla_{[\rho}\mathbf{F}_{\mu\nu]} &= 0 \end{aligned} \tag{26}$$

The fact that matter does not interact with curvature is still only an approximation. One can see why this kind of "coexistence" is not unification. Suspicions of adhocness are

[106] There are several known formulations of this principle, including some attributed to Einstein: inertia and gravity are manifestation of the same underlying structure, without being identical (Janssen 2005, 58). From this and from the results of the special theory of relativity it necessarily follows that the symmetrical fundamental tensor *g* determines the metrical properties of space, the inertial behavior of bodies in it, as well as gravitational action. My formulation is approximate. See (Norton 1985 233; Einstein and Grossmann 1914 224) for a history of this principle.

[107] For a more comprehensive discussion of this principle see (Wüthrich 2006)

.

well founded in this case because a duality between matter and field lingers here. This is not unification at all, but a strategy to rule out a possible influence of **EM** over **GR** (Norton 1985).

8.1. Differences between GR and EM

A somewhat simplistic way to prepare unification is to assess how different or how similar two theories are. Similarities are necessary to unification, but they are nothing more than necessary conditions.

There are some evident similarities and dissimilarities between **EM** and **GR**. (Norton 1992, 17-94). First, the two theories are different on several accounts. Unlike **SR** which is intimately connected to **EM**, Maxwell's theory evolved in the 19^{th} century separately from the theory of gravitation. The gravitational and the electromagnetic fields were seemingly *independent*, although reciprocally *consistent*. They were thought of as being independent, because one can imagine worlds having the electromagnetic interactions switched off with gravitational interactions only, or vice versa. The two interactions act independently on "probes", e.g. charged particles and masses and in each and every point of space a gravitational field can act independently and consistently with the electromagnetic force. For the physics of those times, both theories could be developed separately without paying too much attention to the "coupling" between them. But as I mentioned before, there are also some deeper differences between gravitation and electromagnetism and the way they transport energy.

First, during the 19^{th} century, gravity was thought of as the paradigm of "action-at-a-distance", whereas according to Maxwell and—on the contrary to what Hertz thought—**EM** was the paradigm of action by proximity through waves and throughout a medium (“the ether”). For the action-at-distance theorist, the energy resides in masses as potential energy in the sense that it has power to produce effects at a distance without being carried on. This has been the case of theories of gravitation in the 19th century. For the action-through-a-medium, energy is in the field (i.e. in ether or in space) and in the bodies, and has two forms: potential, and dynamical, i.e. kinetic. This was precisely Maxwell’s point when he accepted that there is motion in the ether (Morrison 2000 83)**. EM** seemed to be a theory that described the dynamics of the electromagnetic field in spacetime, depending on the distribution of charges, currents, dipoles, on their motion and on the motion of observers. To all of these, one can add the boundary conditions needed to solve Maxwell equations in their invariant form. Even now it is not clear whether gravitation is an action-at-a-distance interaction or on the contrary it is carried by particles. In **GR**, gravitation is mediated by a field. For the period I focus on here, gravitation was clearly an action-at-distance interaction.

Second, even if the gravitational field is similar to the electric field, there is nothing in gravitation corresponding to the magnetic field. In its textbook formulation, the gravitation field depends neither on the dynamics of the probe particles, nor on the dynamics of the observer. In the case of electromagnetic forces, the dynamics of probe particles determined the values of the magnetic field.

Third, gravitation is exclusively attractive, whereas electromagnetic force can be repulsive or attractive.

Fourth, Einstein realized an old problem that had haunted the theory of gravitation since Newton: gravity cannot be described by linear equations though it has a linear formulation for weak fields which is very useful for a large class of systems (for example, it can predict gravitational waves among other new predictions; obviously it does not apply to intense gravitational fields close to a black hole for example). His **GR** is definitely not a linear theory. The prospect of incorporating gravitation and electromagnetism within one set of equation seemed an even more difficult project.

Fifth, one of the great discoveries of **GR** was that there is no part of spacetime without gravitational potential, unlike **EM** potential.

> If we consider the gravitational field and the electromagnetic field from the standpoint of the ether hypothesis, we find a remarkable difference between the two. There can be no space[,] nor any part of space without gravitational potentials; for these confer upon space its metrical qualities, without which it cannot be imagined at all. The existence of the gravitational field is inseparably bound up with the existence of space. On the other hand a part of space may very well be imagined without an elecctromagnetic field; thus in contrast with the gravitational field, the electromagnetic field seems to be only secondarily linked to the ether, the formal nature of the electromagnetic field being as yet in no way determined by that of gravitational ether. From the present state of theory it looks as if the electromagnetic field, as opposed to the gravitational field, rests upon an entirely new formal motif, as I thought nature might just as well have endowed the gravitational ether with fields of quite another type, for example, with fields of a scalar potential, instead of fields of the electromagnetic type. Einstein's quote from 1920 as in (Renn 2007 618).

Sixth, there is another difference related to the group invariance of the two theories. We know now that the gauge symmetry groups of the two theories are very different. **EM** theory has the symmetry U(1), whereas **GR** has the symmetry group **GL**, i.e. the diffemorphism group Diff(M)—discovered much later than U(1)—and this can constitute a major hindrance for unification. These results were not known during the relativity years, although the **EM** symmetry group was intuited by Weyl.

Lastly, one controversial issue within **EM** is its interaction with matter. The photon model advanced by Einstein was able to explain the thermal equilibrium of matter and radiation.[108] The alternative semiclassical models in which matter is quantized but the field is classical were short lived.[109] If fact, electromagnetic interaction is quantized and has only an approximate description by classical fields.

The interpretative questions raised by **GR** were not simple: its quantization is perhaps the most controversial aspect. But simpler questions can be asked in the case of **GR**: what is this theory about? In the original interpretation, **GR** is a theory of the field metric *g* and its dynamics that depends on the presence of matter, dust, pressure and fields—electromagnetic fields included. The gravitational field transport energy and in fact can take the form of gravitational waves. One source of puzzles was that even in the absence of matter, the field equations describing *g* have non-trivial solutions! If the metric depends on the presence of matter, when matter is not present, where the values of *g* come from? What is the cause having as effects the "wrinkles" in *g*? We cannot use Mach's principle anymore in this case. Another puzzle is that according to **GR** there is no way to take out the *g* field from the spacetime. The *g* field it is a physical field but with unphysical properties—or close to what unphysical is: it acts upon itself, but cannot be acted upon, as Einstein put it in the 1922 book on relativity (Einstein 1955 55). If relativity violates the action-reaction principle according to which physical objects influence and in general are

---

[108] The term "photon" was not used by Einstein. His original term was "Lichtquant".

[109] The Bohr-Kramers-Slater (1924) theory is the final attempt to describe **EM** fields by classical equations and matter as quantized.

influenced by others, it becomes contrary to the scientific thinking for Einstein; a paper of 1924, quoted in (Brown 2005 140). All these interpretative questions were more or less amplified by the question on the coexistence between **EM** and **GR**.

Other interpretative issues related to **GR** were also not clear at that time. No general covariant conservation law of the energy momentum of the *g* existed. In addition, for simple distribution of matter, the generally covariant field equations could not uniquely determine the field. This meant a failure of physical causality and would lead to the "hole argument": in other words, if *g* is completely determined by the energy-momentum tensor, then the coordinate system cannot be arbitrarily chosen, so the covariance breaks (Earman and Norton 1987, 515-525; Butterfield 1989, 1-28; Belot 1996, S80-S88).

One trivial attitude toward the unification of **EM** and **GR** is to deny the possibility of unification because of these severe differences and interpretative issues on both sides. This is wrong on at least one account. In the case of previous successful attempts to unification, the classes of unified phenomena had seemed beforehand terribly different. Think of the answer of a Neo-Aristotelian scholar to Newton's unification of terrestrial and celestial phenomena. She would balk at the major differences between the sizes of planets or satellites compared to the size of rocks and cannonballs, to the colossal distances or velocities of these bodies compared to the small ones we deal with on Earth, etc. Even the trajectories of these objects belong to two different classes: trajectories of celestial objects are all periodic, whereas terrestrial objects fall toward the center of Earth very quickly. It is easy to see why such differences are merely apparent. Now we know that there is no major fundamental difference between the two classes of phenomena and their trajectories are simply various types of conics. The same can be said about the unification of electric and magnetic fields or other unrelated phenomena: for instance, aurora borealis, friction, lighting, light, chemical reactions are at the fundamental level all electromagnetic phenomena, although they look absolutely different. Maybe gravitation and electromagnetism are only apparently different, although in reality identical. In his Leiden address (1920), Einstein echoed this hope: "Of course, it would be a great advance if we could succeed in comprehending the gravitational field and the electromagnetic field together as one unified conformation." (Renn 2007 619; Einstein 1920).

In short, interpretative issues related to **GR** or **EM** did not hinder the attempts to unify them. There was a hope that unification could partially solve some of them. When interpretative issues are present in two theories, unification could solve some of them. We will see in the Kaluza-Klein theory how unification acted as a problem solver, for at least some of the problems of EM and GR discussed above.

8.2. Similarities between GR and EM

If so, how reasonable is it to think that a theory that describes the gravitational field can be unified with a theory that describes an electromagnetic field? Let us discuss the other argument for unification of gravitation and electromagnetism. Various intuitions favoring such an unification: both theories deal with particles and fields, both theories are formulated as **PDE** of the second order, both aim to be invariant under spacetime coordinate transformations, although under transformations with different groups.

In Maudlin's account of unification of forces, a necessary condition is that the two interactions share the same dynamics (Maudlin 1996).[110] There are also some striking similarities and analogies between the laws expressing the attraction or repulsion of the

[110] I discuss at large Maudlin's conditions in Section 12.6.

gravitational and electric fields: the form of the partial differential equations in **EM** and **GR** and the fact that both theories can be inferred from a variational principle.

Poisson equations. We need to live with the fact that **GR** is a non-linear theory. But there is an idealization of it which is strikingly similar to **EM**, called the linear approximation of **GR**. Gravity and electrostatic forces have the same expression: they are both inverse proportional to the square of the distance between the sources (masses or charges). Coulomb's law of interaction between two charges q_1 and q_2 is:

$$\mathbf{F} = \frac{q_1 q_2}{4\pi\varepsilon_0}\frac{\mathbf{r}}{r^3} \tag{27}$$

where ε_0 is the permeability of the vacuum, is very similar to Newton's law of attraction between two masses m_1 and m_2 (G is Newton's universal constant):

$$\mathbf{F} = G\frac{m_1 m_2}{r^3}\mathbf{r} \tag{28}$$

In addition, both forces can act at any distance r. All fields having this form are described by a Poisson equation. For a Newtonian potential Φ, the Poisson equation is:

$$\nabla^2\Phi = 4\pi G\mu \tag{29}$$

(μ is the mass density and G is the constant of universal attraction). For an electric potential *V*, the Poisson equation is a direct consequence of Maxwell equations:[111]

$$\nabla^2 V = -\frac{\rho}{\varepsilon_0} \tag{30}$$

According to some interpretations, Einstein had grounded his "new" relativity theory on some analogies between the gravitation field and the electric field. In his 1920 Leiden Lecture Einstein stated that: "the space-time theory and the kinematics of the spe-

[111] Here ρ is the density of electrical charge.

cial theory of relativity were modeled on the Maxwell-Lorentz theory of the electromagnetic field. This theory therefore satisfies the conditions of the special theory of relativity, but when viewed from the latter it acquires a novel aspect." (Einstein 1920).

As of the advent of **GR**, **EM** was already a developed and complex mathematical theory, so he naturally sought a common ground with such an advanced theory. Maxwell equations successfully describe how electric and magnetic *fields* respond to charges and currents. Einstein's field equations were intended to show how *the metric* $g_{\mu\nu}$ responds to the presence of energy and momentum so Einstein started to envision g as a field by an analogy with the **EM** theory. However, the mere analogy with **EM** and Newtonian gravity was insufficient for developing **GR**. It can be said that in fact it was one of the ending points. Einstein tried to infer a theory of gravitation that at the limit would have the Poisson equation form.[112]

There were major dissimilarities in **GR** from **SR**. The formalism had to move from the scalar "talk" of potential fields like Φ and V to the tensor "talk" of $g_{\mu\nu}$ and $F_{\mu\nu}$. In this sense, the parallel between the two tensors was heartening. However, for this shift Einstein needed the help of a mathematician.[113] In Einstein's formalism the scalar equation (31) has to be rewritten as an equation between tensors because both the curvature of the metric and the energy and momentum are not simply scalars or vectors. An analogical reasoning can suggest a field equation having a form like (31) and (30):

$$[\nabla^2 g]_{\mu\nu} \propto T_{\mu\nu} \tag{31}$$

[112] See the English translation of "Die Grundlage der allgemeinen Relativitätstheorie" (1916) in (Lorentz and others 1952 111-164), where Einstein explicitly demands of the theory of relativity to obey the Poisson equation.
[113] Legend has it that in the early-1910s Einstein told M. Grossmann "Grossmann, you must help me or else I'll go crazy" (Pais 1982 212).

Indeed, from a formal point of view, this equation preserves the differential form of a Poisson equation. The second derivative of the metric is proportional to the energy-momentum tensor. The most important task was to find the operator on the left hand side of it. Finding a good candidate for $[\nabla^2 g]_{\mu\nu}$ was not a breeze. Einstein relied heavily on some of his collaborators' knowledge and M. Grossmann was the first one to come up with a solution.

Fortunately, there are some candidates for the awkward operator $[\nabla^2 g]_{\mu\nu}$. Unfortunately, there are too many, too. Here the first drive is to take the d'Alembertian operator on tensors $\nabla^2 = \nabla^\mu \nabla_\nu$. In a Riemannian metric this has been proven to be zero, because all the covariant derivatives of $g_{\mu\nu}$ vanish: i.e. $\nabla_\rho g_{\mu\nu} = 0$.

8.3. Gravity and electromagnetism on a different par: Einstein Field Equations

The analogy with the **EM** equations breaks here. What baffled Einstein for several years was that Riemannian metric seemed too strong a constraint for the yet-to-be-born theory.

Even if physics had not helped much, this time mathematics provided a solution to the problem. Being familiar with differential geometry, Grossmann chose the Riemann tensor $R^\rho_{\sigma\mu\nu}$ as part of $[\nabla^2 g]_{\mu\nu}$ for geometrical reasons (Pais 1982 212-217). This tensor is analogous to a second derivative for scalar functions because it is defined using the Christoffel symbols and their derivatives. Christoffel symbols are related to the concept of displacement of vectors. They have many explanatory virtues in general relativity: in a

Riemannian, non-flat spacetime the displacement of a vector changes its components but preserve their length. Christoffel symbols encode this change. If in a system of coordinate x^ν a vector $\mathbf{A}^\nu$ at a point P is displaced in a neighbor point P' with coordinates $x^\nu + dx^\nu$, then the value at P' is $\mathbf{A}^\nu + \delta\mathbf{A}^\nu$ where $\delta\mathbf{A}^\nu = -\Gamma^\nu_{\alpha\beta}\mathbf{A}^\alpha dx_\beta$. The quantity $\Gamma^\nu_{\alpha\beta}$, called the Christoffel symbol, gives the amount by which the component ν of the original vector depends on its own component α when it is displaced on the direction β with an infinitely small displacement *dx*. If all Christoffel symbols vanish, there is locally a Minkowski metric and all $\mathbf{A}^\nu$ conserve their orientation during all possible parallel transport. Christoffel symbols are also used in the definition of the covariant derivative defined as:

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\sigma} V^\sigma \tag{32}$$

used extensively in differential geometry. Given a metric $g_{\mu\nu}$:[114]

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2} g^{\lambda\sigma}(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}) \tag{33}$$

which is a non-linear combination of $g_{\mu\nu}$ and its first order derivatives. This expression appeared in (Einstein and Grossmann 1914, 225) as Grossmann had been familiar with the results of Christoffel, Riemann, F. Klein, Ricci and Levi-Civita in differential invariants, which were absolutely necessary in order to find the right form of $[\nabla^2 g]_{\mu\nu}$ in (32).

The Riemann tensor is a second form of the metric defined as:[115]

$$R^\lambda_{\mu\nu} = \partial_\lambda \Gamma^\lambda_{\nu\mu} - \partial_\nu \Gamma^\lambda_{\lambda\mu} + \Gamma^\rho_{\lambda\sigma}\Gamma^\sigma_{\nu\sigma} - \Gamma^\rho_{\nu\sigma}\Gamma^\sigma_{\lambda\mu} \tag{34}$$

---

[114] See Appendix B for more up-to-date discussion of the concept of metric.

[115] "Christoffel symbols" are notations and *not* tensors because they don't act like tensors. I use here the obvious notation $\dfrac{\partial g_{\mu\nu}}{\partial x^\rho} = \partial_\rho g_{\mu\nu}$. Alternatively, there is the "comma notation" $\dfrac{\partial g_{\mu\nu}}{\partial x^\rho} = g_{\mu\nu,\rho}$ common in the older literature.

The most relevant contractions of the Riemann tensor is the Ricci tensor:

$$R_{\mu\nu} = R^{\lambda}_{\mu\lambda\nu} \tag{35}$$

and the Ricci scalar:

$$R = g^{\mu\nu} R_{\mu\nu} \tag{36}$$

The *saga* of how Einstein and Grossmann arrived at the form of the Einstein Field Equation (**EFE**) is more complicated. Many times Einstein took the wrong path and many times he guessed the solutions. In several instances he ignored relevant facts and spent too much time on details. In 1913 Einstein and Grossmann turned away from the study of the Ricci tensor and then gave up the idea of a generally covariant equation (Norton 1984). They dropped and then came back to Ricci tensors several times. They tackled the idea of energy as well as the Newtonian limit from several perspectives; either denying them or taking the major constraint to the **GR**. I am more interested in the final result of this toiling: the known form of the Einstein field equations, inferred only in 1916:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu} \tag{37}$$

which has the desired form (32). Here $G_{\mu\nu}$ is the Einstein tensor and it is a simple shortcut for the quantity $R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$.[116] It is common to write Einstein Field Equations (**EFE**) as $G_{\mu\nu} \propto T_{\mu\nu}$ or schematically:[117]

---

[116] The constant G (a.k.a Newton's constant) in (30) and the tensor $G_{\mu\nu}$ in (38) have no connection whatsoever.

[117] A different result unbeknowst to Einstein or to Hilbert was the Bianchi identities $\nabla^{\mu} G_{\mu\nu} = 0$ from which one can infer the conservation law $\nabla^{\mu} T_{\mu\nu} = 0$. He correctly derived the conservation law in October 1916. This omission made Hilbert to believe that electromagnetism is the consequence of gravitation. This result is not correct. The story of the Bianchi identities is discussed in (Pais 1982 275-277).

$$\begin{pmatrix} \text{variations of a field} \\ \text{second order PDE in:} \\ F_{\mu\nu}, g_{\mu\nu} \end{pmatrix} (=) \begin{pmatrix} \text{a measure of a source} \\ \text{linear terms in:} \\ T_{\mu\nu}, J_{\mu} \end{pmatrix} \tag{38}$$

Now we see that the metric of the general theory of gravity has been designed such that it satisfies the type of equation as the electromagnetic and Newtonian field. The “differential line element” s^2 which expresses the length between two points infinitesimally displaced by dx^μ is generalized from its Minkowski form in **SR**:[118]

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu \tag{39}$$

to:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \tag{40}$$

By this a correspondence between **SR** and **GR** is granted. This is again in accord with the acclaimed principle of correspondence according to which Einstein’s equations should correspond at the limit to Newton’s formula. After some uneasy months of mathematical struggles, Einstein was able to find the formulation of **GR** which perfectly corresponds to Newtonian mechanics.[119]

8.4. A uniform description of EM and GR: the Einstein-Hilbert action

We see that g is expressed as a function of T, which encodes energy and momentum carried among others by the electromagnetic field. This is not unification, but another way of restating that g depends on the distribution of energy and matter. Hilbert took another step towards a more *uniform* treatment of **GR** and **EM**. Compared to Einstein’s treatment,

---

[118] $\eta^{\mu\nu}$ is the Minkowski metric with the signature (- + + +).

[119] Another problem of “correspondence” in the case of **GR** was its correspondence with classical mechanics in order to capture the full “explanatory” and “prediction” store of Newtonian mechanics.

Hilbert is closer to what one can dub “unification”. Hilbert thought in terms of action and independently of Einstein provided a different justification of **EFE**.[120]

Hilbert started from the formalism of **CM** which admits a very elegant and simple formulation in terms of Lagrange density and action. For simple systems, the Lagrangian $\mathfrak{L}(q_i, \dot{q}_i, t)$ is the difference between the kinetic energy and the potential energy and for a large class of systems is a function of generalized positions q_i and their time derivatives $\dot{q}_i$. Variants of the action principle had been used in analytical mechanics under various names since Maupertuis (1742) and Euler (1746), the most well-known being the “principle of stationary action” (*aka* Hamilton’s principle). In the Lagrange formalism of analytical mechanics, the system is described in a configuration space Q of independent generalized coordinates q_i in which the most relevant feature is the Lagrange function $\mathfrak{L}(q_i, \dot{q}_i, t)$. Lagrangian mechanics of a system can be derived from a variational principle by taking the action integral of the Lagrangian.

[20] ***(Principle of stationary action)****: If the system of the problem is described in configuration space by a Lagrangian* $\mathfrak{L}(q_i, \dot{q}_i, t)$*, then the action integral:*

$$\mathrm{S} = \int_{t_0}^{t_1} \mathfrak{L}(q_i, \dot{q}_i, t)\, dt \qquad (41)$$

is stationary: $\delta \mathrm{S} = 0$.

From this principle one can infer the Euler-Lagrange equation of motion (for all q(t) and all $t_0 < t < t_1$):

[120] Strictly speaking, Hilbert’s proof had been published five days before Einstein presented his equation to the Prussian Academy on November 20, 1915.

$$\frac{d}{dt}\frac{\partial \mathfrak{L}}{\partial \dot{q}_i} - \frac{\partial \mathfrak{L}}{\partial q_i} = 0 \tag{42}$$

The same can be inferred in areas other than analytical mechanics. In a classical theory of fields, for all sets of classical fields Φ^{i} for which the Lagrangian "density" is $\mathfrak{L}(\Phi^i, \nabla_\mu \Phi^i)$, the variational principle states that the action integral is stationary. The action has been used in **SR** to derive the equation of motion, in **EM** to infer Maxwell equations, and in quantum field theory to derive Dirac equations. Although it is true that for almost all these theories there is a non-variational method available to formulate the equations of motion, for my purposes the action principle is a powerful formalism that can methodologically unify many physical theories. The principle of action can be generalized to the concept of optimality that is used in biology, economics etc. It is a different problem whether the variational principle can be extended from a mathematical method to reality. Does the variational principle have a reference in the world?

*Hilbert's derivation of the **EFE***. In **GR** there are two "alternative" formulations to the standard derivation of the field equations: the Lagrangian and the Hamiltonian. For many physicists, the Hamiltonian formulation is extremely important in the context of quantum gravity.[121] But giving the generality and simplicity of the Lagrangian formulation, physicists think that it "contributes further to the aesthetic appeal of general relativity" (Wald 1984 450).

In the Lagrangian **GR** one wants to find an action integral invariant under spacetime transformations. In the same vein as the action principle, we want to find an action

[121] The Hamiltonian formulation also reveals the way Einstein's equation describes the evolution of spatial metric in time. Canonical quantization of gravity uses the Hamiltonian formulation. See (Wald 1984, 491; Wüthrich 2006; Wüthrich 2005, 777-788)

integral for a field ϕ (it can be any field: a scalar, a vector or a tensor field) over a finite region Ω of spacetime:

$$S(\Omega) = \int_{\Omega} \mathfrak{L}(\phi, \partial_{\mu}\phi, x) d^{(4)}x \tag{43}$$

that remains invariant under a spacetime transformations:

$$\overline{x}^{\mu} \rightarrow x^{\mu} + \xi^{\mu}(x^{\nu}) \tag{44}$$

The Lagrangian that leaves the action integral (44) invariant, i.e. δS(Ω)=0 for all (45) is called "invariant density". A similar reasoning was used in deriving Einstein equations from an action by Hilbert as early as 1915. The problem of choosing ϕ is as serious as the choice of a derivative of the metric in Einstein's method. According to Hilbert, the simplest choice is to look for a scalar ϕ. Hilbert's scalar density is: $\mathfrak{L} = \sqrt{-g}R$, i.e. the Hilbert type of action that depends only on the absolute value of g and on its first order derivatives $\partial_{\rho} g^{\mu\nu}$:

$$\mathrm{S}_{Hilbert} = \int R\sqrt{-|g|}d^4x \tag{45}$$

By applying [20], we take the action to be invariant to small variations. The variation of g is $\delta g = |g| g^{\mu\nu} \delta g_{\mu\nu}$ and the variation of $g_{\mu\nu}$ is:

$$\delta g_{\mu\nu} = \partial_{\mu}\xi^{\rho} g_{\rho\nu} + \partial_{\nu}\xi^{\rho} g_{\mu\rho} + \xi^{\rho}\partial_{\rho} g_{\mu\nu} \tag{46}$$

Such that $\delta\sqrt{-g} = -\frac{1}{2\sqrt{-g}}\delta g = \frac{1}{2}\sqrt{-g}(g^{\mu\nu}\delta g_{\mu\nu}) = -\frac{1}{2}\sqrt{-g}(g_{\mu\nu}\delta g^{\mu\nu})$. The same can be done for δR, although the computation is longer. As:

$$R^{\rho}{}_{\sigma\mu\nu} = \partial_{\mu}\Gamma^{\rho}_{\nu\sigma} - \partial_{\nu}\Gamma^{\rho}_{\mu\sigma} + \Gamma^{\rho}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} - \Gamma^{\rho}_{\nu\lambda}\Gamma^{\lambda}_{\mu\sigma}$$

we get:

$$\delta R^{\rho}{}_{\sigma\mu\nu} = \partial_{\mu}\delta\Gamma^{\rho}_{\nu\sigma} - \partial_{\nu}\delta\Gamma^{\rho}_{\mu\sigma} + \delta\Gamma^{\rho}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} + \Gamma^{\rho}_{\mu\lambda}\delta\Gamma^{\lambda}_{\nu\sigma} - \delta\Gamma^{\rho}_{\nu\lambda}\Gamma^{\lambda}_{\mu\sigma} - \Gamma^{\rho}_{\nu\lambda}\delta\Gamma^{\lambda}_{\mu\sigma}$$

$\delta R_{\mu\nu} \equiv \delta R^{\rho}{}_{\mu\rho\nu} = \nabla_{\rho}(\delta\Gamma^{\rho}_{\nu\mu}) - \nabla_{\nu}(\delta\Gamma^{\rho}_{\rho\mu})$ and after some algebra:

$$\delta R = R_{\mu\nu}\delta g^{\mu\nu} + g^{\mu\nu}\delta R_{\mu\nu} = R_{\mu\nu}\delta g^{\mu\nu} + \nabla_{\sigma}\left(g^{\mu\nu}\delta\Gamma^{\sigma}_{\nu\mu} - g^{\mu\sigma}\delta\Gamma^{\rho}_{\rho\mu}\right)$$

One can use Stokes theorem to show that the boundary of the last total derivative does not contribute at the final integral. Finally, $\frac{\delta R}{\delta g^{\mu\nu}} = R_{\mu\nu}$. For the precise calculations of δR, see (Carroll 2004 161-4).

By setting $\delta S_{Hilbert}=0$ and the "vacuum condition" $T_{\mu\nu}=0$, there are three sources for the variation of S_H:

$$\delta S_H = \delta S_1 + \delta S_2 + \delta S_3 \tag{47}$$

where:

$$\begin{cases} \delta S_1 = \int d^4x\sqrt{|g|}g^{\mu\nu}\delta R_{\mu\nu} \\ \delta S_2 = \int d^4x\sqrt{|g|}R_{\mu\nu}\,\delta g^{\mu\nu} \\ \delta S_3 = \int d^4x\sqrt{|g|}R\,\delta\sqrt{|g|} \end{cases} \tag{48}$$

Starting from some boundary condition considerations, one can explain why S_1 does not contribute to δS. Keeping in mind that $\delta\sqrt{|g|} = -\frac{1}{2}\sqrt{|g|}g_{\mu\nu}\delta g^{\mu\nu}$, we can bring δS_3 to the form of δS_2. The equation of vacuum can be recovered from (49) and it has the expected form of (38) with $T^{\mu\nu}=0$. Vacuum solutions are important, but what if matter or electromagnetic fields are present? Hilbert proposed to add to S_H the term reflecting the presence of matter S_{matter}. The variation of "matter" action is taken to be connected to the stress energy tensor by a simple relation:

$$\delta S_{matter} = -\frac{1}{2}\sqrt{-g}\, T_{\mu\nu}\, \delta g^{\mu\nu} \tag{49}$$

By applying the variation of action to $S=S_{Hilbert}+S_{matter}$, one can recover exactly the form (38).

Hilbert wrote the first action for **GR** in 1915. Later on, Einstein used the same principle to derive the equation of motion for a static field. He showed that from $\delta\int ds = 0$ and from the metric (41), that the equation of motion can be inferred. Within weeks of Hilbert's paper going to press, Einstein published a revised version in which he inferred the equations without any reference to systems of coordinates and without appealing to "material phenomena" (Lorentz and others 1952; Norton 1984 150; Einstein 1916).[122]

For unification, the Lagrangian method has some advantages. First, **EM** theory admitted such a formulation and the same type of formulation for **GR** allows for an easy unification with **EM** or other classical fields. Second, the Lagrangian also clearly identifies a natural candidate for the source term coupling the metric to matter fields. Third, by applying Noether's theorem, the action allows for the discovery of conserved quantities through the symmetries of the action.

Hilbert claimed that his action is reminiscent of **CM**: "Also here it is seen—as was shown for the usual relativity theory by Planck—that the equations of analytical mechanics have a significance which far exceeds that of Newtonian mechanics" (Pais 1982 203)**.** If this is correct, action and its minimization have the function to link relativity to **CM**.

[122] Historians debate Einstein's primacy and honesty in deriving the equations as well as on the intellectual property between Einstein and Hilbert. In 1997 some archival work made by Leo Corry ruled out the possibility that Einstein plagiarized Hilbert. It seems Hilbert's draft paper still had some mistakes corrected in the published version of 1916. The discovery of the equation occurred in November 1915 and Einstein arguably had them first although it was only a matter of weeks. (Corry, Renn, and Stachel 1997) provides copious historical details of this controversy.

Unfortunately, the principle of least action has a very limited application to quantum systems. But for the purposes of classical field theory, it can arguably play a major role in unification.

Despite its simplicity and universality, in [20] the action and the Lagrangian density were "guessed". Similar to Einstein's and Grossmann's initial discovery, using an action principle implies some guesswork. But this time the inference is closer to the practice in other areas such as electromagnetism, classical mechanics and fluid dynamics where the action principle reigns.

Even if **GR** can be formulated without it, the action principle constitutes a major step forward in providing a simpler formulation based on the analogy with **CM**. Of course, nothing could be inferred about the existence or the reality of the action. I do not promote the principle of least action to a unificatory element, although many physicists at the turn of the 20th century thought so: Philip Frank, Max Planck, Otto Hahn and even Hilbert himself were altogether mesmerized by the principle of least action "unificatory" power.[123]

I take the principle of least action as a step toward a more uniform treatment of **EM** and **GR**. I think that the principle does not act as an unification element here, although it is crucial as a method of a unitary treatment of the two interactions. It is a schema of inference that can be applied to various theories but it lacks any explanatory power. Virtually we can "unify" almost anything by applying it. It is nothing more than reasoning based on an analogy of the form of the Lagrangian in **GR**, **SR** and **EM**. It can be approached it in the context of Kitcher's account as a stringent argument (maybe even too stringent). It is not at

[123] For a comprehensive discussion on the principle of least action, see (Stöltzner 2003). The importance of this principle for the contemporary physics is still under scrutiny.

all clear what the status of this principle is. Is it a regulatory principle or is it a computational tool? For any given set of equations (Einstein field equations, Klein-Gordon, Dirac, etc.) a suitable action can be defined such that from the variational principle the equations of motion can be inferred. The action S seems in this case a simple mathematical artifact. The "guessing" of the S_{matter} by analogy illustrates this operation well enough.

The EM action. There is a sense in which EM is disunifed in the Lagrange formulation of the **GR**. There is no electromagnetic action S_{EM} to be added to the gravitational action S_H. The only way to do this is through the stress-energy tensor. It seems that electromagnetic field cannot be treated on the same footage as its contribution to the action is added through T. Indeed, even in a non-variational deduction of the **EFE**, the key concept that helped Einstein to encompass electromagnetism and matter was T. If only electromagnetic sources are present, then:

$$T_{\mu\nu} = F_{\mu\rho}F_{\nu}^{\rho} - \frac{1}{4}g_{\mu\nu}F_{\rho\sigma}F^{\rho\sigma} \tag{50}$$

In the Unified Field Theory approach, encoding the electromagnetic fields within *T* does not constitute properly speaking unification because the matter and the electromagnetic field in the stress-energy tensor are not described by the same equation as the metric *g*: They affect *g* rather than being described by the same formalism. The ideal situation is one in which we incorporate *g* and F within the same mathematical structure. The form of the Einstein field equations illustrates the separation between the field and the matter or energy.

8.5. Weyl's unification and his non-Riemannian metric

Hopes that unification could go further than what **GR** offered were high in the second decade of the 20th century. In the 1920s, some physicists including H. Weyl, E. Cartan, A. A. Eddington, etc., had aimed to the unification of **GR** and **EM**. Einstein's theory of general relativity is pseudo-Riemannian because the metric has the form (41), irrespective of fields and matter. Riemannian geometry is based on a postulate: in a Riemannian manifold, the infinitesimal parallel transport of a vector around a closed curve changes its orientation, but not its length, on returning to an initial point.[124]

Cartan's classification. One of the first attempts to find the right geometry for **GR** is due to Élie Cartan. The term affine connection was used by Weyl in his "Raum, Zeit, Materie" (1918) in direct connection with the **GR**. In a series of papers from 1923 to 1925, Cartan tried to provide a rigorous classification of affine spaces and a general framework in which projective geometry can be reconciled with differential geometry (Cartan 1923; Cartan 1924; Cartan 1925). In Cartan's schema, both the classical mechanics and Einstein's theory occupied a specific place. The same can be said about Eddington's and Weyl's attempt to unification. After Eddington's and Schrödinger's results in the affine connections, the relation between affine connections and general relativity was more manifest than before. Cartan associated general relativity to the affine connection. According to Cartan, the characterization of affine spaces can be given by estimating two integrals. The first one is given by the transport of an arbitrary vector. The simpler case described by Cartan is the parallel transport of a vector and of a system of coordinates along a closed

[124] I suspect that a full understanding of the geometrization procedure needs a short digression into the field of differential geometry. I present in Appendix B several mathematical results in differential geometry.

curve in a manifold. For the sake of the argument, we can perform this on a subspace embedded in manifold. If we do this on a plane in the $\mathbb{R}^3$ manifold, two quantities will be always conserved at the end of the parallel transport: the length of the vector and the orientation of the system of coordinates. On the contrary, if we perform the same operation on a sphere in $\mathbb{R}^3$, only the length of the vector will be conserved. The orientation of the coordinate system is not preserved in general.[125]

In a more rigorous way, we can associate three types of curvatures to an affine connection. The rotation curvature is defined as:

$$\Omega_{\mu}^{\nu} = -R_{\mu\rho\sigma}^{\nu} dA^{\rho\sigma} \tag{50}$$

and the homothetic curvature:[126]

$$\Omega = \Omega_{\mu}^{\mu} = -R_{\mu\rho\sigma}^{\mu} dA^{\rho\sigma} \tag{50}$$

The torsion is:

$$\Omega^{\rho} = -(\Gamma_{\mu\nu}^{\rho} - \Gamma_{\nu\mu}^{\rho}) dA^{\mu\nu} \tag{50}$$

where $A^{\mu\nu}$ is the area of the closed loop.

The Euclidean space has all these curvatures vanish. A torsion-free space will preserve the length of the vector, while a null curvature (both rotation and homothetic)

---

[125] Let us suppose you leave the North Pole on the following route: North Pole, New York, London, North Pole and you travel in straight lines. In this case, the orientation of the coordinate system is not preserved, but the length of the vector is preserved. Nonetheless, if you travel in a straight line from North Pole to the South Pole and back in a straight line, the length and the orientation are preserved.

[126] Homothetic transformation dilates distances with respect to a fixed point A.

preserves the orientation. I synthesize Cartan's classification in this table: See in general (Cartan 1923, 325-412; Tonnelat 1965).

Table 1: Types of geometries in Weyl and Cartan

4D geometry	Rotation curvature	Homothetic curvature	Torsion
Euclidean space: Newtonian mechanics	0	0	0
Riemanian space: **GR** (Einstein, 1915)	any	0	0
Affine space: Weyl, Eddington	any	any	0
Affine space with torsion: Einstein (1945), Schrödinger (1947)	any	any	any

The major difference between Weyl and Cartan and Einstein is that for Weyl the metric is not anymore Riemannian when the **EM** fields are present. In the standard theory, the Riemannian geometry is replaced by the "affine connection", albeit the topological features of the Minkowski manifold are preserved. In "Gravitation und Elektricität" (1918), Weyl stated that this is a remnant of the Euclidean geometry. Why should we preserve the length of a vector through parallel transport?[127] His "pure infinitesimal geometry" has been conceived as a genuine local geometry in which length is also transported and changed from one point to the other. In his geometry, the comparison between lengths that are not path-independent is like the comparison of a vector's orientation in Riemannian geometry. Weyl's geometry is weaker than Riemann's in the sense that it is only conformal—only the angle between vectors is preserved by their parallel transport along a closed curve, neither their length nor their orientation. In such a general structure only a subset of transformations is preserved. Consequently, g is no longer a universal

[127] Translated in (Lorentz and others 1952 201-216).

quantity or at least it becomes something else in the presence of electromagnetic field. But does the **EM** field change the *type* of connection of spacetime?

In Einstein's theory, the ten independent magnitudes present within $g_{\mu\nu}$ are all potentials of the gravitational field and their values depend on the surrounding physical magnitudes of mass energy and momentum. The Newtonian concept of the magnitude of gravitational force is replaced in this version of geometrization by "degrees" of curvature of space-time. For example, Earth's mass determines a curvature in the space-time fabric that manifests to an observer as a source of gravitational action. It is fundamentally a source of nothing, rather it is merely a local deformation of geometry. The free falling objects in this model do not follow the "pull" of a force, but simply the "laziest" track along the "bumps and hollows" of space-time.[128]

Here is the meaning of geometrization in Weyl: a physical field can be considered geometrized if its potential is to be found as part of the metric of the theory (Pasini 1988 291). A purely "geometrized" object is entirely constructed from the Riemann curvature tensor which is derived in turn from $g_{\mu\nu}$. Weyl takes into consideration a class of conformably equivalent metrics [g] on which there is a linear and torsion-free connection such that: $\nabla g = -2\mathbf{A} \otimes g$ where $\nabla_{\lambda} g_{\mu\nu} = -2A_{\lambda} g_{\mu\nu}$ and $\mathbf{A} = A_{\mu} dx^{\mu}$ is a differential 1-form. By taking two points P and Q and a curve γ that unites them, the length l(Q) of a vector X in point Q measured with a representative $g \in [g]$ is:

$$l(\mathrm{Q}) = \exp\left(-\int_{\gamma} \mathbf{A}\right) l(\mathrm{P}) \tag{51}$$

The lengths are the same if and only if the curl of **A** vanishes, i.e.:

[128] These metaphors are from (Ryckman 2005 219).

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = 0 \tag{52}$$

The most natural way to integrate electromagnetism in the theory of gravitation is to identify the F in (56) with the electric strength tensor $F_{\mu\nu}$. (O'Raifeartaigh and Straumann 2000 3).

In the absence of any **EM** field, the metric becomes Riemannian, the length is not anymore path dependent and the class of conformal metrics [g] is reduced to one element g. This illustrates the perturbative unification of **GR** which I alluded to in this section. The **EM** field is a measure of how perturbed the metric is from the Riemannian form. It is clear in this case that **EM** is part of the covariant derivatives of *g*. It is also the first step towards the conceptualization of gauge theories that will a play crucial role decades later in the genesis of the Standard Model of particles.

I synthesize here Weyl's major results:

- Weyl provided a new form of action which is gauge-invariant to the group transformation U(1), (unlike Einstein-Hilbert action (46): $S(g,A) = -\frac{1}{4}\int Tr(\Omega \wedge *\Omega)$ where Ω is the curvature form and $*\Omega$ its Hodge dual.[129] The action *S(g,A)* splits in a curvature part and in a "electrical" part because:

  $$Tr(\Omega \wedge *\Omega) = Tr(\hat{\Omega} \wedge \Omega) + F \wedge *F$$

  where $\hat{\Omega}$ is the metric piece.

[129] The dual of an (orthonormal) basis of p-vectors is a (n-p)-vector obtained by "wedging" together all the basis 1-vectors not appearing in the p-vector, then multiplying by the norm of that p-vector. For a Minkowski 3-space with the signature $ds^2 = dx^2 + dy^2 - dt^2$, here are the Hodge duals: $*(dx \wedge dy) = dt$; $*(dy \wedge dt) = -dx$; $*(dt \wedge dx) = -dy$; etc . See (Misner, Thorne, and Wheeler 1973 Ch. 15)

- The conservation of electric charge as a gauge invariance;
- the gauge transformation**.** A change of calibration ("gauge") in the metric $\overline{g} \to e^{2\lambda} g$ induces a gauge transformation in A: $\overline{\mathbf{A}} \to \mathbf{A} - d\lambda$. This result will be discussed later in the context of Klein's unification.

The major problem with Weyl's unification is its unrealistic empirical consequences. Einstein and Pauli quickly replied to Weyl's 1918 paper by complaining that his theory does not apply to the real world. We live in a simpler world than Weyl thought. If Weyl was right, then the atomic spectra of two atoms of hydrogen would change as a function of the path they had followed. In other words, the emission of an atom should depend on its "history", a feature that has never been observed for any known elementary particle. In the same manner, the length of rods and clocks should be path-dependent. It seems that we do not leave in a non-Riemannian physical universe: most probably we live in a pseudo-Riemannian world. In Einstein's and Pauli's eyes, these were sufficient reasons to reject forever Weyl's theory (Ryckman 2005 81, 86).

Since Weyl, the attempts made by Einstein and by others to generalize the Riemannian metric or to throw out completely the tensor calculus have not been very promising. It is true that the metric itself does not deviate from its Riemannian form by the mere presence of **EM** field or matter. However, the idea that there are gauge quantities that can determine the variation of a field by this mechanism is crucial in the gauge theories.

Chapter 9. Geometrization of physical forces

Despite Weyl's interesting result, the main trend in relativity was to keep the Riemannian metric: it has been considered the "natural" metric, instantiated in the physical world. So it is not surprising that from the beginning physicists (Kaluza and Klein included) aimed to formulate their theory by using the Riemannian metric.[130] In the unified field theories it is central to develop a uniform way of relating forces to the spacetime structure. Informally, if we want to unify the physical field Φ_1 with the physical field Φ_2 we want to have the same type of relations between the spacetime structure M and the fields Φ_1 and Φ_2. If they relate differently to the spacetime one can ask whether they should be unified after all. This is why the possibility of a unification of physical forces is intrinsically related to an older philosophical idea, the geometrization program—a reductionist program in mathematical physics according to which dynamics can be described by a geometrical structure. If we can geometrize all physical interactions, unification is possible. I suggest here that geometrization is a possible path to unification. Although not necessarily the only one, it plays an essential role in the Kaluza-Klein unification.

At the end of the 19th century there was a strong dichotomy within physics. On the one hand, space had been described by *geometry* because experience cannot teach us anything about mutual parts of space. Experience is about bodies in space, not about space itself. This became a controversial issue at the beginning of the 20th century. There were several types of geometries available at the end of the 19th century, classified according to their group of transformation (the Lie group being the most preeminent). Which one is the truest

[130] In this dissertation I focus only on Riemannian metric so I leave the topic of non-Riemannian metrics.

one? Can we decide on an empirical basis which geometry is the physical geometry? Poincaré and a host of other philosophers and mathematicians answered this question in the negative. For S. Lie (1890), F Klein (1871), and H. Poincaré (1887) at the end of the 19th century geometry was the study of a group of transformations of idealized, rigid bodies and not the study of space itself. Poincaré also postulated that from these possible groups we need to choose the standard (*étalon*) to which we shall refer natural phenomena. Space itself as a framework for sensations and representation must not be taken as an *a priori* form of intuition, *pace* Kant (Torretti 1984 ch. 3; Magnani 2001 ch. 5; Poincaré 1920). On the other hand, physical fields were described by *physics* in which laws of nature can be expressed in partial differential equations (**PDE**). But all **PDE**s refer explicitly to space and to time. These equations are invariant under some transformations (for example the Galilean group of transformations). Spacetime had its own symmetries, whereas the evolution of systems in spacetime had its dynamical symmetry. It was like the science of spacetime had nothing to do with the science of objects in spacetime. Are physical space and time related to the space and time the geometers spoke of? Mechanics postulates space and time, but without describing them. It was like space is described by a different science than mechanics.

Several options have been available during the history of physics—needless to say each of these options can be interpreted in a strong, reductionist way or in a weaker way:

- *(A)* Spacetime is physical, i.e. there is a kinematics and a dynamics of spacetime structure and spacetime is similar to a physical field;
- *(B)* Spacetime has no influence over and no interaction with the dynamics of the physical bodies;

(C) Spacetime is not part of reality and it is not necessary to describe the dynamics of physical bodies. Reality is constituted by bodies and physical fields only.

(D) Spacetime acts upon the physical bodies and influence their dynamics, but not the other way around;

(E) Geometrization: there is a description of physical interactions such that all physical quantities can be described geometrically; the dynamics of the physical fields is part of a hidden geometrical structure.

The least plausible option for classical physics was (A). Before Newton, Aristotelian physics endorsed (B) or (C) in the sense that bodies had some properties as "absolute rest", "natural places" such that it rendered superfluous spacetime with its geometry. Bodies were carrying with them the spatial or temporal properties we usually attribute to spacetime points. Galileo was able to reject (C) and show that properties such as "absolute rest" and "absolute time" as well as the Galilean symmetries refer to space and time–not to objects.[131] After Newton, the dominant answers to this puzzle were (D) and (E). The end of this story is that in **GR**, Einstein provided a whole theory that endorsed (A).

Who has pursed (E)? It seems that it refers to our representation of the world and not to the world itself. By "geometrical description" I mean a general representation in terms of geometrical structures that include but are not limited to: curves, trajectories, geometrical objects in abstract spaces, volumes, shapes, surfaces and all the mathematics

[131] In a later interpretation of the "Erlangen program" of Felix Klein (1893), the Galilean symmetries are active transformations of spacetime: Spatial translations: $t \to t; \vec{x} \to \vec{x} + \vec{a}$. Time translations: $t \to t + \tau; \vec{x} \to \vec{x}$. Shear mappings: $t \to t; \vec{x} \to \vec{x} + \vec{v}t$. Rotations and reflections: $t \to t; \vec{x} \to \mathbf{R}\vec{x}$, where **R** is an orthogonal matrix.

that comes with them; for a more general discussion on geometrization see (Boi 2004; Kalinowski 1988). Why geometrization and not representation by numbers? If numbers are coordinates, then the answer is that the world does not come equipped with coordinate systems. We can change the coordinates based on convention and convenience. But coordinates are not the only numbers. Quantities are expressed by numbers. Even for such cases, there is a general trend in mathematics to think that numbers do not represent the world (Melia 2000, 455; Field 1980, 130). There are no numbers in the world, strictly speaking. But there are geometrical objects in the world: trajectories, curves, possible paths of particles, and light cones. In this sense, as numbers are not present in the world, we cannot say for example that number one is more real or closer to reality than π or than the number i such that $i^2 = -1$. We can nevertheless argue whether there are planes in the world or solids in six dimensions or if a certain curve can or cannot be the trajectory of a real particle. The substativalist would say that curves, solids, and points are parts of an entity called spacetime and that numbers are representations or conventions referring to it. The relationalist will take objects as real and curves as conventions or mental representations of real or possible relations between objects. In both cases, we do not need to believe in numbers, unless we adopt a Pythagorean philosophy. If the geometrization program is at least in principle possible, dynamics of objects in spacetime can be thought of as geometrical objects. Numbers come later, if they are needed.

Strong geometrization is bad geometrization. In a strong reading, we want to reduce two dynamics of Φ_1 and Φ_2 to the structure of spacetime. There is no dynamics except the geometry of spacetime. This idea is rooted in Descartes, but it was endorsed and popularized in the 1920s by H. Weyl, A. A. Eddington and especially H. Meyerson. Later on, Eddington and Weyl would retract this strong geometrization, so Meyerson is perhaps its only proponent. Descartes thought that geometry and mathematics were the canon of any deduction, including reasoning in physics. Matter has to be reduced to the kinematics of the featureless extension which is space. Descartes' dream of a pure mathematical physics based on property of space and time was in fact impossible and its replacement by Newtonian physics witnesses the failure of the Cartesian geometrization. As a follower of Descartes' *panmathematicism*, Meyerson wrote that the role of scientific explanation and theories is to replace the "infinite diverse world around us by identity in time and space, which clearly can be only space itself" (Meyerson 1991 137). Meyerson went so far to interpret the structure of spacetime as the *Ding an sich* (Ryckman 2005 240; Meyerson 1985 (1925) 212). It is even worse than geometry being summoned from a Platonic heaven to become sovereign over a messy world. This extreme geometrization rooted in metaphysics is so fraught with problems that I take it as the worst geometrization program.

Why is this so bad? First, the metaphysical geometrization negates the diversity and the complexity of the real world in order to impose a simpler although incomplete description. Stating that geometry is rational and the world has to be described geometrically in order to be rational is a bad argument which has a Hegelian flavor. Second, Meyerson's reading of Descartes succumbs finally to an irreconcilable dualism between mind and the sensorial experience.

After being introduced to Meyerson's work, Weyl, Eddington as well as Einstein denied that relativity was a geometrical theory of the world. Einstein explicitly wrote in 1928 in a review of Meyerson book that the essential point of Weyl and Eddington was to show a possible way to *represent* gravitation and electromagnetism under a unified point of view." (quoted in Ryckman 2005 240) Later in a letter to Barret from 1948, Einstein explicitly dismissed the idea that **GR** is geometrizing physics. Field concepts were more important for Einstein, at least in the 1940s (Lehmkuhl 2008 86). For Einstein, geometry was nothing other than a theory of idealized solid bodies. Eddington also dismissed Meyerson's theory by confessing that he and Weyl had not set out to create a geometrical theory of the world, but to "seek the physical reality by approved methods" (Eddington 1920 183).

A different, weaker meaning than Meyerson's has to be sought in which geometrization is a better method of representing the dynamics of physical fields.[132] How do we relate a physical field to the structure of spacetime without the intricacies of Meyerson's theory? And what's the difference between geometrization and representing something under a geometrical point of view? First, let us see what can be geometrized.

Geometrization of other subdomains of physics. Depending on the standard used, almost any discipline within physics can be geometrized. Classical mechanics can be represented in abstract spaces like configuration space or phase space. The dynamics of a system of N particles can be reduced to a trajectory in the abstract space and features of this trajectory provide an accurate representation of its evolution. Statistical mechanics makes use of this geometrization of mechanics and describe the possible evolution of systems as surfaces or volumes in such an abstract space. A similar approach to thermodynamics was developed by C. Carathéodory in 1909 (Brown 2005 136sqq.). In optics, the behavior of light can be geometrized given a specific idealization of electromagnetic wave theory. In this case, geometrization is the result of an idealization that proved useful. Later on,

[132] For a detail of Meyerson account see (Zahar 2007, 291; Ryckman 2005 ch. 9).

quantum mechanics and quantum field theories have admitted several formalisms more or less remotely related to this concept of classical geometrization.

I focus especially on the geometrization of classical mechanics. Newtonian physics is grounded on the explicit assumption that space, time and physical events are independent. In Newtonian physics, spacetime is the background of the motion of physical objects and influence this motion without being influenced by it. As there is no causal influences from physical fields to spacetime, one can say that spacetime is not dynamical and at the limit non-physical. Option (D) is frustrating from a metaphysical point of view because it is reminiscent of something like a "prime mover" or an absolute agent that acts upon the world. An easier solution was to look for a common representation of spacetime and physical objects. As dynamics of spacetime in itself was still too audacious an idea, the only option was to look for a geometrical description of spacetime. Witness that the "world-postulate", "world-line", light-cones, etc and all the lingo in today's physics is the result of Minkowski's geometrization. The "kind of union of space and time" anticipated by Poincaré in 1906 and by Minkowski in 1908 is a geometrization of kinematics of **EM** by using 4-vectors (Poincaré 1906; Minkowski 1909esp. sect. IV)

Dynamical concepts are easily translated into the geometrical language: the motion of a certain reference frame with a velocity v is a rotation in the Minkowski spacetime; accelerated systems follow curves with non-null torsion, energy and momentum form a 4-vector that is related to possible trajectories starting from one point, etc. The 3-D appearances of this beautiful 4-D reality are "complicated" and even ugly. For Minkowski, we are compelled to admit that only in 4-D the "relations [among physical quantities] reveal their inner being in full simplicity", and to realize that these relations on a three-dimensional space "forced upon us *a priori* they cast only a very complicated projection" (quoted inBrown 2005 131; Minkowski 1909, 104-111).

Is gravitation geometrized in **CM**? The answer is in general negative. In many respects geometrization was only partially accomplished before **GR**. Classical gravitation is contingently geometrized, whereas **GR** fully geometrizes gravitation. We can look at the Newtonian gravitational theory from the perspective of Newton's laws. The inertial mass and the gravitational mass are equal, but only *contingently* equal in **CM**. In a flat spacetime the trajectory of a particle in the absence of any potential is a straight line:

$$\frac{d^2x}{dt^2} = 0 \tag{53}$$

The motion of an object in a gravitational field with a potential Φ is characterized by a generalized form of Newton's second law in which the inertial mass and the gravitational mass are in principle different; see also **(30)**:

$$m_i \frac{d^2x}{dt^2} = -m_g \partial\Phi \tag{54}$$

In a flat space, the solution to this equation is a curve, i.e. the trajectory of the particle. Rewriting :

$$m_i \frac{d^2x}{dt^2} + m_g \partial\Phi = 0 \tag{55}$$

is the equation of a line. One can interpret this as the equation of a straight line in a curved spacetime. The curvature is given by $\partial\Phi$ and in **CM** m_i=m_g for all probe particles. But this is still contingent. In this case, all trajectories will have the same standard of being a straight line. This means that inertial frames are the same for all particles, no matter what their masses as long as they obey $m_i = m_g$. Whether this is the case or not remains contingent. If for a given particle the inertial mass is not equal to its gravitational mass, then the definition of straight line would depend on their ratio. This means that there will be different geometries with different definitions of straightness for different ratios.

In general, when a force is present, the acceleration is not null. Even if the two could look the same, in fact in Newtonian gravitation there is a difference between straight trajectories in non-flat spacetime and curved trajectories in flat spacetime. The presence of forces means something more than the curvature spacetime can be endowed with.

More recent analyses of Newton-Cartan's theory, especially Trautman's (Trautman 1965), have revealed interesting insights in the genesis of **GR**; according to D. Malament, Newton-Cartan is significant for showing how close relativity is from the Newtonian theory of gravitation and in what respects "in coordinate-free, geometric language, Newtonian gravitation theory (or, at least, a certain generalized version of it) is the 'classical limit' of general relativity" (Malament 2007 sec. 3.2). The general conclusion of Malament's analysis is that geometrization of gravitation in Newton-Cartan theory is apparent.

9.1. Geometrization of gravitation in GR

In **GR** the previous story about geometrization has a similar meaning although other problems arise. According to one reading of the Equivalence Principle, the empirical observations made on a probe in a *homogeneous* gravitational field are indistinguishable from the observations made on the same probe subjected to a uniform acceleration with the same value. As Newtonian physics, **GR** postulates the ratio between m_i and m_g as being 1. Inertial reference frames defined by geodesics include: (a) objects at rest, (b) in constant motion, or (c) gravitationally accelerated. Accelerations in **GR** are of two types: either grav$_{i}$tation$_{a}$lly induced or not. The first type of acceleration is relative in the sense that it depends on the *syst*em of reference. The non-gravitationally induced accelerations remain absolute in **GR**. According to the Equivalence Principle, **GR** makes no distinction between straight trajectories and "gravitationally-accelerated" trajectories, i.e. trajectories which are bent by gravitation. Gravitation is the curvature of spacetime. There is a difference between **EM**-induced curved trajectories and trajectories induced by gravity. Moreover, Einstein postulated the dynamical character of the spacetime in the sense that he changed its status. A substantivalist would say that he demoted it from a physical field to a geometrical aspect of the spacetime, i.e. the curvature itself. A relationalist about spacetime (such as Einstein), would promote spacetime from geometry to physics.[133]

Here is a definition of geometrization useful to the present discussion:

> ***Def 9*** **Full geometrization of a field**: A scalar field φ or a vector field A_μ is fully geometrized if it appears in the expression of the tensor $g_{\mu\nu}$ and not in that of $T_{\mu\nu}$.

If gravitation is a manifestation of spacetime geometry then we adopt the geometrization picture. What does it mean that gravitation is an aspect of geometry (Wald 1984 67)? Some authors speak of gravitation as a *manifestation* of curvature (Misner, Thorne, and Wheeler 1973, 1279) or of gravitation as *arising* from the spacetime curvature (Hartle 2003, 582). If on the contrary, one wants to call geometry the manifestation of a gravitational field, then one adopts the field interpretation of **GR**.[134] D. Lehmkuhl recently called

---

[133] This idea was presented to me by Jonathan Bain in a private communication. As suggested by Reichenbach, despite the dispute on the status of spacetime, there are two or three positions in respect of the relation between geometry and the theory of relativity. In fact, the two possibilities are logically consistent (Reichenbach 1958 (1928), 295).

[134] R. Feynman and S. Weinberg are maybe the radicals in rejecting the geometrization program.

this interpretation the "gravitization of geometry". According to Lehmkuhl, a third option would be to take the "egalitarian" stance, according to which both interpretations are important. The strong egalitarian position claim that "geometry and gravity are conceptually identified within the theory, making them two names for one and the same 'thing'" (Lehmkuhl 2008 84). A weak "egalitarianism" is the position that even if gravitational field and geometry can always be swapped, they are still different entities. If every mathematical object within the **GR** formalism has both geometric and gravitational significance, then the egalitarian position wins.

This is close to the standard conventionalist interpretation of **GR**. Should we geometrize the gravitational force or not? It seems this is a question of taste and choice difficult to make. Is the description of a flat geometry + gravitational force the same as a curved geometry + no gravitational force? A pure conventionalist would say that they are ultimately the same, as long as we stay away from topics like quantization of gravity, or thermal relativity.[135] The jury is still out. The egalitarian position according to which one can switch from the field interpretation to the geometrical interpretation seems to be a viable alternative to the two radical positions, but it appears contentious. In fact, even in **GR** there are fields that are not gravitational, but geometrical. Matter fields are added to **GR** by the means of $T_{\mu\nu}$. In order to accept the egalitarian position, a different distinction is needed between matter fields and non-matter fields. Adding non-matter fields to the theory complicates the egalitarian position. Lehmkuhl acknowledges this challenge when he discusses the Brans-Dicke theory of gravitation that adds a scalar field similar to the one

[135] Several physicists expressed this conventionalist position, including C. Rovelli. In this specific instance, Reichenbach adopted the view that geometry became an expression of the gravitational field (Rovelli 1997 193).

present in Kaluza-Klein theory. In the context of Kaluza-Klein theory, one may ask the egalitarian whether geometry is “electromagnetized” more than it is “gravitizated” or similarly whether electromagnetic field is geometrized in the same way gravitation is.

Another problem related to the egalitarianism position proposed by Lehmkuhl and this is related to scientific explanation. Let us accept for the sake of the argument that the geometrical and gravitational interpretations can coexist, even in the weak egalitarianism position. I suspect that we would not be able to entertain both interpretations if explanations and predictions are at stake. If I want to explain phenomena, let us say the radiation of a black hole, I need to pick one of the two explanations: the geometrical one or the gravitational one. Reichenbach had a strong preference for the field interpretation based on causality: the field produces certain geometry of spacetime. Or alternatively, once we enlarge the area of research from mere **GR** to cosmology, let us say, we need to ask questions involving primacy: “what is generated from what? Is geometry the effect of gravitation or the other way around?” When we ask similar questions within the quantum gravity, we need to make a choice, too. I suppose that egalitarianism needs to face such challenges. In the case of Kaluza-Klein unification, the preeminent questions I am interested in are related to explanation and secondly to the causal relation between geometry and gravitation. These are unsolved issues in the egalitarian position and I claim that the analysis of the Kaluza-Klein theories can shed light on these convoluted issues in respect of making a choice between geometrization or gravitation. I do not see the pure conventionalist position as tenable either.

9.2. Geometrization of EM in GR

Geometry-gravitation-matter? What if we move beyond **GR** and discussed geometrization in a broader context? What happens when we need to add electromagnetic fields to **GR**? Or matter? We can replace $\partial\Phi$ in (59) by a different term given by the electric potential. This will define the straight line of electromagnetism, won't it? But for a different ratio q/m, we need to define a new straight line. Different particles with different ratios would follow different straight lines. The situation seems worse than in the case of the Newton-Cartan theory: a geometrization of forces depending on an internal parameter such as the charge q is not possible because it leads to inconsistencies.

I see a possible solution: keep the **EM** forces not geometrized and associate them with terms similar to ρ. But electromagnetism is not only a collection of charges, currents and/or densities. It comes with its own vector potential A_μ and with a scalar potential V. Notwithstanding other differences in the nature of gravitation and **EM** interaction, on the LHS of both **(32)** and **(31)** we encounter second-order derivatives of fields depending on matter or charges—codified in the RHS terms.

Among the problems with interpreting $T_{\mu\nu}$ includes considering it as a *source* of curvature. But in relativity the dependence relation in **(32)** does not work only one way: the gravitational field can do work on matter and vice versa. Transforming $T_{\mu\nu}$ into a tensor with trace means creating matter out of fields. However, the whole procedure is convoluted and it was a deadlock. Einstein considered his theory of relativity as a theory only of the gravitational field, independent of the theory about the structure of matter. Quantum mechanics had been expected to give the full apprehension of the dynamics of matter.

Einstein's field equations (**EFE**) already had the **EM** encoded in the right hand term $T_{\mu\nu}$ and there was no way to add the **EM** field in the expression of $g_{\mu\nu}$. The only way is to make $g_{\mu\nu}$ dependent on **EM** by the **EFE**. Gravity was geometrized while **EM** fields were not. In the original form, Einstein had geometrized *only* the gravitational field, whereas all other fields including the electromagnetic field are part of the $T_{\mu\nu}$ and *not* of the metric $g_{\mu\nu}$. Einstein thought of a decomposition of the tensor $T_{\mu\nu}$ into a traceless part that represents massless fields like the electromagnetic one and a trace part that encodes the masses. For a brief duration by the end of 1915 he even tried to resuscitate the electromagnetic program by admitting that matter is a product of the interaction between electromagnetic and gravitational field. Matter was not to be something given primordially and physically elementary.

If matter is to be encoded in the $T_{\mu\nu}$, then where is the place of **EM** fields? The GR answer is: also in the $T_{\mu\nu}$. A possible alternative to the options discussed here is to geometrize it, i.e. to put it in the geometrical part of the **EFE** equations. I focus here on a specific meaning very popular in the 1920s: a physical field had been considered *geometrized* if its potential was to be found exclusively as part of the metric. All other fields, as well as matter and charges, appeared only in the stress-energy tensor $T_{\mu\nu}$ in **(39)**. The "geometrization" program was intended to move all the non-material fields to G and it is well illustrated by the marble-wood metaphor:

> By analogy, think of a magnificent, gnarled tree growing in the middle of a park. Architects have surrounded this grizzled tree with a plaza made of beautiful pieces of the purest marble. The architects have carefully assembled the marble pieces to resemble a dazzling floral pattern with vines and roots emanating from the tree. To paraphrase Mach's principle: The presence of the tree determines the pattern of the marble surrounding it. But Einstein hated this dichotomy between wood, which seemed to be ugly and

> complicated, and marble, which was simple and pure. His dream was to turn the tree into marble; he would have liked to have a plaza completely made of marble, with a beautiful, symmetrical marble statue of a tree at its center (Kaku 1994 99; Einstein 1936, 313-337).

What if *all* fields could be represented together in the LHS term with g and A_μ encoded in the same mathematical structure? In the light of these similarities, both fields could stem from one and the same universal tensor. Metaphorically speaking, Einstein contemplated the possibility of turning the "wood" of T (the matter) into the "marble" of G (the spacetime) in **(38)**. For Einstein, matter was a term that infected the pure and clean structure of G. He intended to geometrize matter by turning wood into marble—that is, to give a completely geometric origin to matter.

In a fully geometrized theory of field there are no field terms in the stress-energy tensor $T_{\mu\nu}$. But this is only the first stage of geometrization. The second stage is to geometrize matter and charges by encoding them in field equations other than $T_{\mu\nu}$. You want to take one step at a time though. At the beginning, Kaluza's and Klein's formalisms were both vacuum theories, insofar their ontology was populated only with fields with no matter fields. We will see that according to an interesting result in mathematical physics we can represent a matter+geometry theory in N dimension as a field only theory in N+1 dimensions.[136] If we endow the 5-D vacuum world with reality, it looks simpler than the 4-D world. Geometrization as a representation of a known matter field occurs in a new "geometrical" way. The **EM** field becomes part of the g field and only matter, charges, currents, dust, etc. are present in the T tensor.

[136] See Campbell-Maagard theory in Section 13.1.

I mentioned that geometrization is an ideal situation for the unified field theories, but it is not the only one. There are cases in which **EM** fields appear in the expression of the metric without being properly speaking unified with g. In this case one can see a *reciprocal dependency* without unification. A well-known model of the interaction between *g* and **EM** is described by the Reissner-Nördstrom equation (Carroll 2004 255). A charged, non-rotating, spherically symmetric massive body of mass M can be described by Einstein Field Equations with a spherical symmetry. The mass M and its charge act like a source of energy momentum tensor. The result, studied by Reissner as early as 1916, is called the Reissner-Nordström spacetime (Reissner 1916, 106):

$$d\tau^2 = \left(1 - \frac{r_s}{r} + \frac{r_Q^2}{r^2}\right) dt^2 - \frac{dr^2}{1 - \frac{r_s}{r} + \frac{r_Q^2}{r^2}} - r^2 d\theta^2 - r^2 \sin^2\theta \, d\varphi^2 \tag{56}$$

where $r_s = \frac{2GM}{c^2}$ is the Schwarzchild radius and $r_Q^2 = \frac{Q^2 G}{4\pi\epsilon_0 c^4}$ is the length-scale of the charge *Q*. The field equation for this case is given by both the **EFEs** for gravity and by the Maxwell equations for the **EM** field. The electromagnetic strength tensor is present in Einstein's equation through the momentum tensor and the metric enters the Maxwell equations. There is a simple reason to believe that the Reissner-Nordström equation does not unify in the sense desired here. Unification cannot be premised on the sheer fact that one field depends on another field and vice versa.

Notwithstanding the similarities between **GR** and **EM** and the past success of unification, attempts at their unification come with no surprises. However, for the reasons discussed, it was unlikely that their unification would take a simple form. Also, physicists were not satisfied with unification within the action or with Weyl's extreme solution. All the unification attempts discussed in this chapter have their own problems. Kaluza's and Klein's are both alternatives to these attempts to unify. As methods of representing the dynamics of various physical systems, all these attempts to geometrization, mostly successful, brought about elegance and simplifying power. We need to step back a little bit from Minkowski's enthusiasm and admit a fundamental fact: geometrization is only a formalism that can be useful and can simplify calculations and reveal, for example, the simplest form of Lorentz transformations. Further questions remain: are geometrizations more than convenient representations? Do they explain? Do they tell us something about the world? For the specific case of Kaluza and Klein, these are questions are to be addressed in Chapter 16.

Chapter 10. Kaluza's five-dimensional theory

As I have showed in the previous chapter, there were several options at hand to unify **GR** and **EM** and several factors that hindered this unification. First, there was the problem of interpreting **EM**: is it a classical field theory described by Maxwell's equations, or on the contrary a theory about the transport of bits of energy (photons)? As we will see, the unification attempts all started by assuming the former: **EM** is a theory about continuous fields described by the covariant tensor field $F_{\mu\nu}$ and not about quanta of energy moving with the speed of light. This is absolutely natural because a quantum theory of electromagnetic fields, i.e. quantum electrodynamics (**QED**) was not available until late 1920s.[137] Later on, under Einstein's influence, some physicists thought that a *classical* field theory can act as a strong alternative and finally replace **QM**. After years of toil and failure, physicists abandoned for decades this line of thought.[138] A more recent strategy adopted by some philosophers is pluralism in admitting that reality is described by several theories. This would entail liberalism in admitting that there are two descriptions of the world, none more fundamental than the other. The behavior of classical systems cannot be inferred from quantum dynamics, or at least not from the standard interpretation of the quantum dynamics. The classical behavior is independent of the quantum behavior. This strategy is frequently adopted in several interpretations of **QM**. Moreover, we start from the presupposition that both dynamics are inexact. Belot suggested that we can adopt

[137] P.M. Dirac, W. Pauli, F. Weisskopf, etc. were the first to apply quantum mechanics to fields circa 1927. Richard Feynman, Freeman Dyson, Julian Schwinger, and Sin-Itiro Tomonaga developed **QED** in 1940s.

[138] The Bell inequalities and the Kochen-Specker theorem rule out a large class of theories that try to combine quantum and classical descriptions. In a more general context the correspondence principle can be questioned on several bases.

Toulmin's metaphor of maps: each map is more or less accurate but none is the ultimate map: "classical mechanics doesn't say less than quantum mechanics, it says different things" (Dickson 2006; Belot 2000). Classical mechanics is about macroscopic objects, quantum mechanics is about small size objects, better described by wave functions than trajectories. In the absence of a quantum theory of gravity or a quantum description of spacetime, we can entertain the classical, i.e. non-quantized description of reality. Although incomplete, the whole subsequent discussion on Kaluza-Klein makes sense at least. There are not enough reasons to be fundamentalist in respect of **QM** (Belot 2000, S454-S465). There are no serious reasons to think that the unification of **EM** and **GR** is fundamental. But this does not rule it out completely as a viable and serious theory about the world. As we will see, QM played a special role in Klein's theory.

Classical theories are not free of interpretational problems. There are difficulties in interpreting **GR** itself: is $T_{\mu\nu}$ the source of the dynamics of $g_{\mu\nu}$ or on the contrary the latter acts as a source for its own change that adds to the contribution of $T_{\mu\nu}$? Is **EM** just another carrier for energy or it is more than that? What is matter: a special solution to the field equations, or, on the contrary, does it exist independently? Is the Riemannian metric the right one to describe spacetime *everywhere*? Is the structure of spacetime independent of matter and fields?

All these questions had a direct impact on the unificatory programs, Kaluza-Klein included; I focus here on the last question. A material inconsistency among the desire to unify **EM** and **GR** and the mathematical and physical constraints imposed upon them arose in the second decade of the 20th century. Two concepts play a central role in the formulation of **GR:** (a) the metric g and (b) the manifold $\mathcal{M}$.

(a) The first constraint in **GR** is the Riemannian metric which is an invariant of the metric. In **GR** the line element ds^2 **(41)** is conserved after any type of transformation of coordinates. We saw that, unlike a handful of mathematicians (Weyl, Eddington, i.a.), few were happy to give it up. Kaluza and Klein both assumed that the metric *g* is pseudo-Riemannian. In its plain meaning, the Riemannian metric explains how real objects are transported in space. For Riemann and for Helmholtz there was strong evidence that the orientation of objects can be changed by following a different path to return to the same starting point, but not their lengths or angles. Bodies are reoriented after travelling in space, but not stretched, shrunk or deformed in any way. Riemann wrote down the simplest metric that accounted for these properties. Four-vectors should be displaced following the same rule. If a vector A^v at a point $P(x^v)$ in a certain system is displaced to a neighboring point $P'(x^v+dx^v)$, then the value at P' is $A^\nu - \Gamma^\nu_{\alpha\beta} A^\alpha dx^\beta$. $\Gamma^\nu_{\alpha\beta}$ gives the amount by which the ν component of the original vector depends on its own component α when it is displaced on the direction β with an infinitesimal displacement *dx*. Weyl's geometry is weaker than Riemann's in the sense that it is only conformal (only the angle between vectors is preserved by their parallel transport along a closed curve, not their lengths, nor their orientation). But in Riemannian geometry, angles and lengths are conserved in a parallel transport of a vector.

(b) **GR** is restricted to a manifold with a dimension of D=4. This manifold had its strong physical meaning, notwithstanding more recent worries that we live in a 3-D+1 world instead of the 4-D world. In fact, we observe that we live in a

> spacetime with three spatial dimensions to which one dimension (time) can be added in the **SR** formulation of mechanics (and it turned out to be very useful). Unlike **GR**, in the Kaluza-Klein theory, the dimensionality of $\mathcal{M}$ plays a central role in unifying, explaining, and to some extent, predicting phenomena. In few words, Kaluza and Klein strived to show that the dimensionality of the manifold $\mathcal{M}$ is not absolutely given and that we can imagine at least in principle a manifold with more than four dimensions. Unlike other attempts to justify extra spatial dimensions, Kaluza's and Klein's theories are both grounded on the idea of unification. As I will discuss extensively in Part III, for them the manifold unifies and explains. We saw that the properties of the manifold and the properties of the metric define the structure of the **GR**. Yet a deeper incompatibility between $\mathcal{M}$ and g on one hand and the claim of a **GR**-**EM** unification on the other hand arises: they may have different dimensions.

Here are the main assumptions needed for a unified program that claims to be similar enough to **GR**:

[21] *The metric g is (pseudo)-Riemannian;*

[22] *The manifold $\mathcal{M}$ is four-dimensional (4-D)*[139]

and its main claim:

[23] ***GR*** *and* ***EM*** *can be unified by "geometrization" within $\mathcal{M}$ and g in the sense of Def 8 (p. 162).*

Informally, given [21] and [22], [23] was not possible because there is not enough structure to embed **EM** theory in $\mathcal{M}^{(4)}$ with a Riemannian metric g. It seems that the ma-

---

[139] When it is possible, I use the exponent in parenthesis $\mathcal{M}^{(4)}$ to indicate the dimensionality of the manifold.

thematical structure of $\mathcal{M}^{(4)}$ with a Riemannian metric cannot accommodate electromagnetism. The easiest way to see why is to look at the Christoffel symbols of *g*. This is the route taken by Kaluza. As we will see, Klein will follow a different path to unification.

G. Nordström. Before discussing Kaluza's idea, it is relevant to mention a similar approach due to G. Nordström (1914) who attempted to express the metric as a 5x5 matrix that incorporated the electromagnetic field altogether (Appelquist, Chodos, and Freund 1987 50-56).[140] Nordström added another spatial dimension to the existing three in order to obtain an Abelian five-vector gauge field for which a Maxwell-like equation can be written, including a conserved 5-D current. The fifth component of the five-vector potential was identified with gravity while the first four components of the 5D vector potential were components of A_μ. He was the first to explicitly infer from the equations of his scalar field (called f) in 5D that "we are entitled to regard the four-dimensional space-time as a surface in a five-dimensional world." (Appelquist, Chodos, and Freund 1987 53). Like in Kaluza, in Nordström the fifth axis is a "special" direction because all partial derivatives of the scalar field are set to zero, unlike fields on the other three dimensions. Besides Einstein, few reacted to Nordström's result, whereas Kaluza's had a direct impact on subsequent developments.

Another major drawback of Nordström's attempt is that it is based on the electromagnetic program which is a reductive program and not an unificatory program. Unfortunately, gravitation simply does not have the same gauge group as electromagnetism and cannot be reduced to electromagnetism (see p. 321 for details). Nordström's program

[140] I do not intend to discuss in details the role and the importance of Nordström's paper, less known than Kaluza's and unfortunately having only a slight impact on the scientific community, although Einstein commented extensively and praised Nordström in his (Einstein 1913, 487-500).

is yet another failed reduction and not a unification. Both Kaluza and Klein are less reductive and more unificatory in their nature so Nordström's theory is only apparently related to Klauza.

10.1. Geometrization in Kaluza and its historical background

In 1918-1919 Theodor Kaluza took one of the first steps on the road to unification through geometrization. A gifted mathematician from the University of Königsberg, he was deeply influenced by the Kantian philosophy according to which nature is harmonious and can be described by the mathematical language.[141] In 1919 he wrote to Einstein about a bold possibility to ascertain [23] by assuming [21] and by rejecting the "commonsensical" [22]. Kaluza assumed that the manifold $\mathcal{M}$ is five-dimensional and hypothesized on the structure of the new coordinate or parameter of the world. Kaluza did not call the new coordinate a dimension, but a "world-parameter". Nevertheless, he explicitly talks about a mapping of the spacetime onto $\mathbb{R}^{(5)}$. Einstein replied: "The idea... of a five dimensional cylinder world would never have dawned on me... at a first glance I like your idea enormously... the formal unity of your theory is startling" (Pais 2000 331). During his stay in Berlin, Einstein communicated Kaluza's paper to the Prussian Academy after about two and a half years (December 1921). One reason is that the journal of the Prussian Academy, *Sitzungsberichte der königschen Preussichen Akademie der Wissenschaften zu Berlin*, had limited the submissions to eight pages because of the paper shortage after the war. Kaluza

[141] Kaluza worked at the University of Könisberg till 1929 when he moved to Kiel and then moved to Göttingen where he was appointed professor. He died in 1954. After 1921, he has never published articles in physics. Short biographical notes can be found in (Vizgin 1994 149sqq.; Wunsch 2005 sect. 15.2). The most important thing to remember is that by being formed in Könisberg and working in Göttingen, Kaluza was intellectually influenced by Kant and had contact with David Hilbert and Hermann Minkowski.

tried to massively shorten his article and that operation took a while. The other possible reason is more subtle. Einstein had an indecisive attitude toward Kaluza's 5-D formalism because of his own convictions close to the logical positivism of those years. For Einstein in the early 1920s, spacetime structure and manifold was totally determined by observational entities: rods, clocks, light rays, and trajectories. A "bare", extended, fourth spatial dimension has not been observed. From 1919 to 1921, Kaluza and Einstein had some mail exchanges in which Einstein praised Kaluza's idea but also expressed his reticence (especially on a detail related to geodesics in 5-D). Einstein was still wavering between Weyl's and Kaluza's attempts to unification, when he concluded in a postcard of October 14, 1921 that the latter "seems in any case to have more to [unification] than [Weyl's]" (Sabbata and Schmutzer 1982 447-457).

The paper "Zum Unitätsproblem der Physik" published in 1921 had a small impact if any upon the physics' community.[142] Besides Einstein and Grommer (Einstein and Grommer 1923, 1-4) who criticized Kaluza, there were almost no immediate reactions to Kaluza's paper until Klein's paper in 1926. Later on, Klein in "Quantum Theory and Five Dimensional Theory of Relativity" (1926), Einstein and P. Bergmann in "On a generalization of Kaluza's theory of electricity" (1938) would revise and interpret Kaluza's paper. Interestingly enough, only in two years (1926-1927) there was a burst of publications on the 5-D theories: O. Klein, F. Fock, F. Mandel, P. Gamow, W. Gordon, P. Ehrenfest, Uhlenbeck, etc. dealt with the theory. Another burst came in the 1931-1933 with the interest in

[142] See (Kaluza 1921) and translations in English in (Appelquist, Chodos, and Freund 1987; Sabbata and Schmutzer 1982; O'Raifeartaigh 1997). I change Kaluza's original notation to fit those in Klein, Einstein as well as in van Dongen and most of the literature on String Theory. The quotes hereby are my own translation of the original.

projective geometry of O. Veblen, B. Hoffmann, D. Schouten, van Danzig, W. Pauli, Einstein, Mayer, etc. and in the 1940s and 1950s by P. Bergmann, Einstein, P. Jordan, Ludwig, Thiry, Rumer, Ikeda etc. in a generalized connections, pseudo-vectors, etc. (Vizgin 1994 160).

I hypothesize that under the influence of Kant and post-Kantian philosophy, Kaluza thought geometrically about the **EM** fields. The only support of my hypothesis here is found in the biographical note due to D. Wünsch (Wunsch 2005; Wunsch and Goenner 2005).[143] This can be a result of Minkowski's and Hilbert's influences, but more remotely Kant's. For Kant, space is the object of geometry. Space is also the basis of our formal intuitions, directly determined from a mathematical point of view.[144] Kaluza believed that all interactions in nature originate from one entity which could be represented in a rational way. He hypothesized that the structure of spacetime should explain the **EM** field, as it explains gravity, which was an idea inspired by Einstein, who succinctly expressed it later: "it often appears to me that the magnetic field of the earth is based upon an as yet unknown connection between gravitation and electromagnetism but I cannot come out of the inconsistencies [Widerspruchen]" (2/27/1925 letter to Kaluza, in (Sabbata and Schmutzer 1982 457)**.** As we will see, Kaluza's theory exemplifies simplicity and parsimony because only matter and electrical charges, if any, are present in the stress-energy tensor *T*. Maxwell's equation as well as Einstein's field equations fall from the 5-D metric as a "natural

[143] A new monography on Kaluza came aut of press right after this chapter was written and was not available to me till mid 2009. See (Wuensch 2007, 716, [32]).

[144] There is a short note in the *Critique of Pure Reason* (§26, B161, note a) that suggested that geometry is the model of knowledge. Being under Hilbert's influence, it is possible that Kaluza accepted Hilbert's critique of Kant. According to Hilbert, we start from the intuitions of space and we proceed to their logic. Hilbert's axioms in the *Grundlagen der Geometrie* (1899) are not a mere analysis of spatial intuitions.

consequence”. Kaluza was not concerned about the presence of matter and charge, as he wanted to accommodate **EM** fields within g. Inspired by the “marble and wood” metaphor, Kaluza devised his theory as a thought that the universe is, strictly speaking, empty of matter and the only real entity is g. But Kaluza was aware that in order to test his theory, a charged probe, i.e. a small test particle with a certain electrical charge, was necessary.

As a field theory, Kaluza’s formalism attempted to unify structures of fields without sources by embedding them into the geometry of spacetime. By this “geometrization”, the fields become aspects of the same entity, the metric tensor (called the “universal tensor” by Kaluza), such that geometry and physics are no longer distinct ways of describing the world (Weingard 1991, nd; Kaluza 1921 859).

Kaluza’s paper is arguably one of the first cases in the history of science wherein a new theory was developed solely to pursue unification as a theoretical virtue. Unification was a virtue sought in itself, not a feature merely accompanying an explanation or prediction of new phenomena. In its intention, Kaluza’s paper was more metaphysical than computational or empirical as it aimed to remove the duality of gravity and electricity, “while not lessening the theory’s [of gravity] enthralling beauty” by directly envisaging the simplicity and the beauty of the theory (Kaluza 1921 859). Kaluza abstained from making new predictions or seeking novel explanations. Kaluza did not intend to provide a self-standing theory to account for new empirical data. Almost all the consequences of his theory are derivable from **EM** and **GR**. Kaluza heavily borrowed from the **GR** conceptual machinery. One of the main aims of his theory was to infer the form of the known geodesics of charged particles. A blatant exception is the presence of the extra scalar field (ϕ

field) that would be a source of both troubles and new ideas in the decades to come. Kaluza speculated on its interpretation in a very informal way.

In order to geometrize the electromagnetic field, Kaluza preserved the Riemannian metric of the manifold and endorsed [21] as, like the majority of the physicists back in the heyday of Relativity, he believed that the Riemannian metric is the "natural one". His assumption was more radical because it envisaged changing the dimensionality of $\mathcal{M}$ instead of the form of the metric $g_{\mu\nu}$. Even if Kaluza's idea is not essentially different from Einstein's formulation and even though it is connected to Nordström's, Weyl's and Eddington's theories, it was the best result at hand from the point of view of unification. If one adds an extra dimension to the manifold, unification in the sense of "geometrization" would be possible and the contradiction among [21], [22] and [23] is removed.

The theory is based on an "unmistakable formal correspondence in the buildup of gravitational and electromagnetic equations",[145] (Kaluza 1921 859).

The vacuum of Kaluza's theory. Kaluza's starting points are:

- The "vacuum hypothesis" (no matter, no charges present):

$$\mathbf{VACUUM}: T_{\mu\nu} = 0 \tag{57}$$

- The Einstein equation for the 4-D vacuum is:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 0 \tag{58}$$

which implies also that for vacuum both R and $R_{\mu\nu}$ vanish:

$$R_{\mu\nu} = 0 \tag{59}$$

- The Maxwell equation of the 4-D vacuum is:

[145] *unverkennbare formale Entsprechen*.

$$\partial_\nu F^{\mu\nu} = 0 \tag{60}$$

As for now, the two theories are presented in their *vacua* form, i.e. without sources of gravitation, electric or magnetic fields. Both theories have multiple classes of non-trivial solutions, even in the absence of fields. As mentioned, this was the main serious problem of a "field interpretation" to **GR**. The field *g* depends on its own dynamics, not only on the distribution of matter and fields. Analyzing the vacuum formulation of these two theories is neither a trivialization, nor a hapless simplification. The aforementioned inconsistency is in fact the inconsistency of two types of *vacua*: the **GR** vacuum and the **EM** vacuum cannot exist together in the sense that they do not have non-trivial solutions.

The 5-D metric and Christoffel symbols. What is the cause of the inconsistency in the system of the statements [21], [22], and [23]? Given (61), where is the place for the **EM** field? The intuitive answer is: somewhere in the expression of *g* itself. But even if both theories have their own vacuum solutions, if one tries to describe gravitation and electromagnetism by the same equations, the perturbation of gravitation due to electromagnetism cannot be qualified anymore as "gravitational vacuum". A four-dimensional world with a Riemannian metric is not enough for both *g* and *F* tensors.

The Riemannian geometry of a 4-D manifold is "saturated" with gravity and there is no place for other interaction because the Christoffel symbols are defined as first derivatives of a single field (Pasini 1988 292). In 4-D there is no way to add the field tensor $F_{\mu\nu}$ to the Christoffel symbols to preserve their properties and later on to impose $R_{\mu\nu}$=0. Christoffel symbols are defined only up to the first derivatives of a single field and they represent the "displacement" of a vector. The "geometrization" of $F_{\mu\nu}$ is not possible in a *four-dimensional Riemannian* manifold $\mathcal{M}^{(4)}$ and this was enough to prove the incompatibility between [21], [23] on one hand, and [22] on the other hand. So the marriage between the gravitational field and the 4-D Riemannian geometry was worn out. The implicit suggestion was to "call a fifth dimension to the rescue" and to reject [22] (Kaluza 1921 860).

The rejection of [22] provided Kaluza with plenty of options with respect to the form of the Riemanian 5-D metric $\hat{g}_{mn}$.[146] Kaluza intended to deduce both **EM** field A_μ and the metric from one and the same tensor. He added to $g_{\mu\nu}$ one row and one column such that electromagnetic potentials are "geometrized" in the metric g and they are *not* sent away in the right hand of Einstein's field equation **(38)** as Einstein had originally proposed. The Riemannian-like metric gives the line-element according to the same formula, generalized to five dimensions:

$$d\hat{s}^2 = g^{(5)}_{mn} dx^m dx^n \triangleq \hat{g}_{mn} dx^m dx^n \tag{61}$$

By invoking a fifth dimension, Kaluza managed to express the **EM** field as part of the metric g. Now, there is room for **EM** within the 5-D $g^{(5)}_{mn}$.

The 5-D Christoffel symbols:

$$\begin{cases} -2\hat{\Gamma}_{4\mu\nu} = \partial_4 g_{\mu\nu} + \partial_\mu g_{\nu 4} - \partial_\nu g_{4\mu} \\ -2\hat{\Gamma}_{\mu\nu 4} = \partial_\mu g_{\nu 4} + \partial_\nu g_{4\mu} - \partial_4 g_{\mu\nu} \end{cases} \tag{62}$$

[146] Hereby I take m, n, o, etc. to be coefficients from 0 to 4 spanning over five dimensions and all hatted quantities are 5-dimensional vectors or tensors. Unlike Kaluza's notation, the fifth dimension is x^4 and the time dimension (t) is x^0. I tacitly corrected some errors from the original paper and from the English translation in (Appelquist, Chodos, and Freund 1987). Throughout my paper, Latin indices refer to manifolds with more than four dimensions (D>4). Greek indices refer only to 4D manifolds and they run from 0 to 3. In the literature the notation is not consistent. Duff for example uses Greek hatted indices for the 5-D manifold (Duff 1995, 22-35). Here I will be more lenient with my notation. $\hat{x}^m = (x^\mu, y), \mu = 0,1,2,3$ is a vector in five dimensions. I use y to any dimensions above the four of the spacetime manifold. For Kaluza and Klein's papers from 1926, y is x^4. Because the presence of upper indices in the expression of *g*, I will use for the 5-D metric two notations: $g^{(5)}_{mn}$ and $\hat{g}_{mn}$. When the exponent (5) is missing and the indices are all Latin, the fifth dimension spacetime is assumed.

help interpreting the $g_{4\mu}$ quantities. The extension of the **4-D** dimension theory to **5-D** is "minimal" in the sense that all the expressions of tensors and the relations between them, as well as the Christoffel symbols are generalized from D=4 to D=5 (Overduin and Wesson 1997 5).

Kaluza hypothesized a formal similarity between the above forms of Christoffel symbols in 5-D and the 4-D expressions of $g_{\mu\nu}$ and $F_{\mu\nu}$. All the expressions of tensors and the relations between them, as well as the Christoffel symbols, are generalized from four to five dimensions:

$$\begin{cases} \text{Second type: } \hat{\Gamma}^{i}_{rs} = \dfrac{1}{2}\hat{g}^{il}\left(\partial_s \hat{g}_{lr} + \partial_r \hat{g}_{ls} - \partial_l \hat{g}_{rs}\right) \\ \text{First type: } \begin{bmatrix} m & n \\ & r \end{bmatrix} = \hat{\Gamma}_{mnr} = -\dfrac{1}{2}\left(\partial_m \hat{g}_{nr} + \partial_n \hat{g}_{rm} - \partial_r \hat{g}_{mn}\right) \end{cases} \tag{63}$$

Kaluza separated the electric influence of $F_{\mu\nu}$ and the gravitational influence of $\partial_\sigma g_{\mu\nu}$ in different "areas" of the Christoffel symbol matrix (the symbols can be arranged in a $5\times5\times5$ matrix form). It is clear that the 4-D part is kept in accordance with the Christoffel symbols:

$$\Gamma^{\lambda}_{\mu\nu} = \frac{1}{2}g^{\lambda\beta}\left(\partial_\nu g_{\beta\mu} + \partial_\mu g_{\beta\nu} - \partial_\beta g_{\mu\nu}\right) \tag{64}$$

The calculations rendered $\Gamma^{4}_{44} = 0$ as expected. After imposing these conditions over the Christoffel symbols, Kaluza derived the form of $\hat{g}_{mn}$. This yields to $g_{4\mu} = 2\alpha A_\mu$, $g_{\nu 4} = 2\alpha A_\nu$ and $g_{44} = \phi$:

$$g^{(5)}_{mn} = \left(\begin{array}{cc} g_{\mu\nu} = \mathbf{G} \text{ "sector"} & g_{4\nu} = \mathbf{EM} \text{ "sector"} \\ \hline g_{\nu 4} = \mathbf{EM} \text{ "sector"} & g_{44} = \phi \end{array} \right) \text{or } \hat{g}_{mn} \tag{65}$$

Where we impose by symmetry $g_{\nu 4} = g_{4\nu}$. Finally, the metric in 5-D:

$$g_{mn}^{(5)} = \begin{pmatrix} g_{\mu\nu} & \alpha A_{\mu} \\ \alpha A_{\nu} & \phi \end{pmatrix} \text{or } \hat{g}_{mn} \tag{66}$$

came "modularized", or separated into several sectors: the gravitational sector $g_{\mu\nu}$, the electro-magnetical sector A_{μ}, and the new, un-interpreted scalar ϕ. One can see that: (a) there is a neat separation between the 4-D gravitation and the 4-D electromagnetism in this expression and (b) g_{mn} incorporates both **EM** and **GR** contributions. This metric later became the prototype of the Kaluza-Klein metric. What is important for the economy of the present analysis is whether this constitutes unification or not (see Chapter 13).

10.2. Logical structure of Kaluza's argument

The form of the field $g_{mn}^{(5)}$ in (70) is to be inferred by Kaluza's argument that has this concise form:

> in order to achieve [23], if [21] is assumed to be true and a set of constraints and approximations are imposed, then [22] is to be denied.

The set of assumptions in Kaluza's argument are:

[24] *All fields are smooth on x^4 ("Cylindricity" of x^4);*

[25] *The **EM** quantities can be identified with parts of the g_{mn} tensor;*

[26] *The fifth dimension is **spacelike**, i.e. the signature is* $(- \; + \; + \; + \; +)$

[27] *The fifth dimension is isomorphic to a real line.*

Kaluza made two simplifying assumptions: he assumed that in the absence of matter the metric is not strongly perturbed from its Lorentzian value; so his theory is a perturbation theory with only the first order effects retained. Moreover, he assumed that probes are non-relativistic. Both assumptions are relatively realistic for a vacuum theory in which only fields are present. Here are the approximations and idealizations of what is his theory:

> ***[28]*** *The metric differs only a little from its Euclidean value ("the weak field approximation");*
>
> ***[29]*** *The velocities of probe particles are small (compared to c).*

Kaluza's result is:

> ***[30]*** *Both Einstein field equation and Maxwell equations can be inferred from one and the same equation;*

10.3. Details of Kaluza's assumptions

Assumption [24] (*Cylindricity of the fifth axis):* From the very start Kaluza tried to include in his formalism the empirical fact that we do not observe the fifth dimension. A surplus structure is present in the 5-D metric and has to be either explained away or set to a definite value. Here is Kaluza's suggestion: in order to take $F_{\mu\nu}$ in, one term out of three is always set to zero in (66) such that $\Gamma_{4\mu\nu}$ and $\Gamma_{\mu\nu 4}$ will contain *only* **EM** terms. The best option is to hypothesize a coordinate system in which the derivatives of all fields on the fifth dimension vanish or approach zero compared to the derivative of the four spacetime directions. This is formally the origin of the "cylinder" condition, the core of Kaluza's unification. Despite its name, the cylindricity condition does not change the topology of the manifold. Cylindricity refers to the values of all physical fields along x^4 and not to its topology.[147]

[147] I come back to this issue in the following chapter in discussion Klein's theory.

The new "direction" x^4 was called a "world parameter" and the first question is why it is hidden to our common measurement procedure. In Kaluza's scarce ontology, the only field present in **5-D** is $g^{(5)}$ and it has zero or small derivative in the direction x^4. Then we experience only three dimensions of space and one of time because fields exist in these four 'directions' which are *not* constant. Small variations of the fields on the fifth dimension mean that the world is "cylindrical" and that the effect of the fifth dimension is of "higher order".

The assumption [24] can be formally expressed as:[148]

$$\mathbf{CYL:}\ \partial_4 g^{(5)}_{mn} = 0 \tag{67}$$

One observation is in order here: (71) is not a generally covariant condition. This can mean two things: that Kaluza's theory does *not* meet two demands: (a) no "prior geometry" and (b) coordinate-independent formulation. This is one of the serious problems with Kaluza, Klein and many other unified field theories.

Kaluza was ready now to take a major step forward and hypothesized the following identification:

$$\mathbf{ID_1:}\begin{cases}\hat{\Gamma}_{\mu\nu 4} = -\alpha(\partial_\nu A_\mu + \partial_\mu A_\nu) \\ \hat{\Gamma}_{4\mu\nu} = \alpha F_{\mu\nu} \\ \hat{\Gamma}_{44\mu} = \partial_\mu \phi\end{cases} \tag{67}$$

where $F_{\mu\nu}$ and A_μ are the **EM** quantities defined in (14), α is a coupling constant and ϕ is an arbitrary scalar field, not yet interpreted. This is premised on the idea that some Christoffel symbols transform like $F_{\mu\nu}$ and this constitutes the major topic of Chapter 13.

[148] ∂_4 is a shortcut for ∂_{x^4} used for typographical reasons.

Assumption [27]: The fifth dimension is non-compact. The fifth axis is isomorphic to $\mathbb{R}^5$ such that the original 4-D manifold is a subspace of $\mathbb{R}^5$.

Assumption [26]: The fifth axis is space-like. The metric has a $(- \quad + \quad + \quad + \quad +)$ signature. From the beginning, extra dimensions were considered spatial, not temporal, which imposes a condition on the values of the field that has to be non-zero and preferably positive.[149]

The “stealthy” scalar field ϕ has for a long time remained without a clear interpretation of the theory as an “arbitrary” parameter (we will later emphasize on its importance). Two Christoffel symbols depend on ϕ and provide an equation for this unknown field:

$$\begin{cases} \Gamma^4_{4\nu} = \dfrac{1}{2}\partial_n g^4_4 = \partial_\nu \varphi \\ \Gamma^\lambda_{44} = -\partial^\lambda \varphi \end{cases} \tag{68}$$

It seems that the tensor calculus imposes the presence of this new surplus structure that is difficult to interpret. One thing was clear to Kaluza: the field φ lives only in the fifth dimension and obeys a specific, Poisson-like equation: $R_{44} = -\Box\phi$. It is still a quantity that has to be interpreted or dropped from the theory. Keeping it uninterpreted burdened Kaluza’s theory with being unphysical: back then theoretical entities needed to have empirical interpretations.

Approximations: weak fields and small velocities (Assumption [28]). Kaluza hypothesized that $g_{\mu\nu}$ differs only slightly from its Euclidian values, i.e. the fifth dimension

[149] A second temporal dimension rises major philosophical and physical problems, although it is not logically impossible. Only in recent years were there attempts to discuss this alternative, and the philosophical investigations are still incipient. Remember that in principle the Campbell theorem allows the embedding in an space with extra temporal dimensions, too.

smoothly perturbs the Euclidian metric. In the "weak field approximation", the metric takes the form:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{69}$$

where $\eta_{\mu\nu}$ is a Minkowskian metric and h is taken such that $|h_{\mu\nu}| \ll 1$. This is in accordance with Einstein's "equivalence principle". In order to provide analytical solutions to the field equations, it is commonly assumed the perturbation formulation **GR** in which the metric differs only a little from its Euclidian value. Kaluza's linearized gravity in **5-D** can be similarly expressed as: $g^{(5)}_{mn} \simeq \eta^{(5)}_{mn}$. One could balk at this assumption because it contradicts the spirit of **GR**. There are some answers here: first, Kaluza's theory is a vacuum theory in which fields are not severely distorted by the presence of matter. Second, a Kaluza theory without the weak field approximation is in fact possible.[150] Third, in **GR** some interesting results can be inferred from this approximation so Kaluza's investigation is worth pursuing.

As a consequence, Kaluza assumed that the third and fourth terms in the Ricci curvature in **5-D**:

$$R^m_{ijk} = \partial_j \Gamma^m_{ik} - \partial_k \Gamma^m_{ij} + \Gamma^n_{ik} \Gamma^m_{nj} - \Gamma^n_{ij} \Gamma^m_{nk} \tag{70}$$

are of the form Γ^2, and since Γ is of first-order, these contribute only to second order effects and can be discarded. In this approximation,

$$R^m_{ijk} \cong \partial_j \Gamma^m_{ik} - \partial_k \Gamma^m_{ij} \tag{71}$$

and it is easy to see that by contracting the Riemann tensor further, the Ricci tensor has a simpler form, too:(Fabbri 2004)(Fabbri 2004)

[150] The result is that the Electromagnetic theory has an non-gauge invariant term (Fabbri 2004).

$$\begin{cases} R_{\mu\nu} = \partial_\lambda \Gamma^\lambda_{\mu\nu} \\ R_{4\nu} = -\alpha \partial^\mu F_{\mu\nu} \\ R_{44} = -\partial_\mu \partial^\mu \phi = -\Box\phi \end{cases} \tag{72}$$

where the separation between index ‘4’ and indices within 0-3 is clearer. Indeed, because of the weak field approximation, the scalar ϕ does not interfere with gravity or electromagnetic field. One can see that the fifth dimension is privileged and this has been since Kaluza one of the major frustrations with theories postulating more than three spatial dimensions.

Kaluza employed another well-known relation between the Christoffel symbols that can be particularized given assumption [24] to:

$$\frac{\partial \Gamma_{4lm}}{\partial n} + \frac{\partial \Gamma_{4mn}}{\partial l} + \frac{\partial \Gamma_{4nl}}{\partial m} = 0 \tag{73}$$

as the 5-D world is empty, i.e. deal only with vacuum solutions. Given (71), both the Ricci scalar and the Ricci tensor vanish:[151]

$$\begin{cases} R_{mn} = 0 \\ R = 0 \end{cases} \tag{74}$$

This approximation allows the dropping of the last two product terms in (75).

Assumption [29]: *The “velocities” on the fifth dimension axis are small*. In order to derive the equations of motions of particles, a second approximation is needed (used in **GR**, too) in which the 5-velocities are such that: $ds^2 \cong d\tau^2$, where τ is the proper time. Kaluza assumed the five-dimension velocity vector $U^m = \dfrac{dx^m}{ds}$ has small components on

[151] The Ricci tensor $R_{ik} = R^j_{ijk}$ is a contraction of the Riemann tensor in 5-D.

the axes x^1, x^2, x^3 and x^4 and a component close to 1 on the x^0. This means that the theory applies only to relatively small velocities *and* to charges of $\rho_0 / \mu_0 \ll 1$ which seems kosher for all practical purposes. This also entails that $d\sigma^2 \simeq ds^2$ and $v^\mu \simeq u^\mu$, where $v^\mu = dx^\mu / d\sigma$. But this second approximation is unsatisfactory for atomic dimensions where u^4 is not at all small for a given density of charge of electron. In this case, [29] is no longer feasible: the electron does not follow a geodesic in $\mathbb{R}^5$ as its u^4 is enormously large. This means that Kaluza's theory would not work for subatomic particles.

10.4. Kaluza's results

*Result [30]: Unification as geometrization of **EM** and **GR***. Finally, it is worth noting that Kaluza's entire theory has an important outcome in the spirit of the desired unification [23]. His result is astonishing: given [24] through [29] and the inferred forms of the Ricci tensor (77), one can derive an Einstein-like field equation in 5-D:

$$\hat{G}_{ij} = \hat{R}_{ij} - \frac{1}{2}\hat{g}_{ij}\hat{R} = 8\pi G\hat{T}_{ij} \tag{75}$$

where T_{ij} is a stress-energy tensor in 5D whose components are the currents **J**: $T_{\mu 4} = J_\mu = \begin{pmatrix} \rho & j_1 & j_2 & j_3 \end{pmatrix}$. Here are the main results inferred from Kaluza's theory:

- Einstein field equations (38) for i,j=0,1,2,3;
- Maxwell equations (14) and (15). In order to infer them, Kaluza used a general relation of Christoffel symbols, true in any dimensions:

$$\partial^j\Gamma^i_{kl} + \partial^l\Gamma^j_{ki} + \partial^i\Gamma^l_{kj} = \partial_k\left(\Gamma^i_{jl} + \Gamma^j_{li} + \Gamma^l_{ij}\right) \tag{76}$$

By taking k=4, and again [24], we infer the homogeneous part of the Maxwell equations. To infer the inhomogeneous equations, from the fifth component of (81):

$$R_{4\nu} - \frac{1}{2} g_{4\nu} R = 8\pi G T_{4\nu} \tag{77}$$

and given [24] we drop the last term with R. Then $\alpha \partial^{\mu} F_{\mu\nu} = 8\pi G J_{\nu}$ which are the inhomogeneous Maxwell equations. Kaluza identified the constant α with $8\pi G$:

$$\alpha = 8\pi G \tag{77}$$

- A Poisson-like equation for the field ϕ: $R_{44} = -\Box\phi$. This leaves again the field ϕ without an interpretation as well as the term R_{55} as its "density". The field is not taken as constant by Kaluza.[152]
- The components of the energy momentum tensor in 5-D: if the perturbative assumption [28] is true, then the Ricci scalar is of higher order in h and the Einstein equations in 5-D are:

$$R_{mn} = \kappa\, T_{mn} \tag{78}$$

and:

$$T_{mn} \simeq T^{mn} = m_0 u^m u^n \tag{79}$$

where m_0 is the rest mass.

- From (77) and the inhomogeneous Maxwell equation (15), one can identify the components of T_{mn} as such that the 4-D energy momentum tensor $T_{\mu\nu}$ is

[152] Contrast this with Klein's assumption. See Section 0.

bordered by vectors representing the currents and densities of charges. It is easy to show that T_{55}=0 and then T_{mn} is:

$$\hat{T}_{mn} = \left(\begin{array}{c|c} T_{\mu\nu}\text{: matter and densities} & J^{\mu} \\ \hline J_{\mu}\text{: currents and charges} & 0 \end{array} \right) \tag{80}$$

and

$$\mathbf{ID_2:}\ \mathbf{T}_{\mu 4} = J_{\mu} \tag{81}$$

where J^m is the 5-D current from (16), such that the force into **5-D** splits in a gravitational part and an electrical part. In the small velocity approximation, it is commonly assumed that:

$$\begin{aligned} ds &\simeq d\lambda \\ v^{\mu} &= \frac{dx^{\mu}}{d\lambda} \simeq u^{\mu} \end{aligned} \tag{82}$$

which allows Kaluza to derive a form of the "density of force" as:

$$\Pi^{l} = \Gamma^{l}_{mn} T^{mn} + F^{l}_{m} J^{m} \tag{83}$$

In the case of charged particles, it is conventionally written:

$$J^{\mu} = \rho v^{\mu} \tag{84}$$

Using the two definitions of the $T_{\mu\nu}$, Kaluza identified u^4 from (90), (88) with a constant proportional to the charge of the particle:

$$\mathbf{ID_3:}\ u^4 = \frac{dx^4}{dt} = \frac{q}{Mc\sqrt{2\kappa}} \tag{85}$$

In brief, here is Kaluza's result: one can infer the fundamental equations of two theories that seemed independent by describing a $g^{(5)}$ in 5-D. One can admit that Kaluza achieved unification in the sense of geometrization and also in the sense of derivation of previous results from one and the same equation. As we will see, there are some important tradeoffs of his unificatory theory and the price is significant.

10.5. Problems with Kaluza's theory: covariation of x^4 and geodesics

At a first sight, Kaluza struggled a lot to gain a little: is this just another way of inferring known results? The price is still very high: his theory is an approximation, does not apply to relativistic particles and it does not describe microscopic and macroscopic objects in the same way. In order to expose the quandaries of this early journey in the extra dimensional theories, I see here several foundational and philosophical problems related to the formalism of the theory.

[31] *Covariation of the metric does not reflect measurability.*

Even before Kaluza's paper was published, Einstein echoed some worries about the empirical evidence for a fifth axis in spacetime (Einstein and Grommer 1923, 1-4). Right after the publication of Kaluza's paper, the main problem Einstein complained about was the general covariance imposed on (65).[153] Einstein associated the covariance of ds^2 with the measurability of a distance or time interval in a locally inertial frame with rods and clocks. But in 5-D, there are no measures of length and duration. "Length" in 5-D does not necessarily have the same meaning as in 4-D. Why should one want to preserve the covariance of ds^2 in 5-D? For Einstein and Grommer, what is at stake is the very condition of "minimal" extension of the **GR** formalism to 5-D. From a physical point of view, the requirement of general covariance of all equations in 5-D continuum is completely at odds with the cylinder condition [24]. In one sense the fifth dimension is special because it has

[153] Einstein & Grommer's paper (Einstein and Grommer 1923, 1-4), published in a rare journal in Jerusalem, is difficult to access. The quotes in this section are from (Vizgin 1994).

the same values of the fields along it, which is not the case with x^0-x^3 and on the other hand it is represented by minimally extending the **GR** theory to it. One can feel a contradiction here. This arises a “dubious asymmetry when one dimension is distinguished from all the others by the cylinder condition, whereas in the structure of the equations all five dimensions must be on equal footing.” (Vizgin 1994 159). The same criticism was expressed by W. Pauli in 1958, in the context of the action principle: why should one choose the curvature as an integrand of the action integral in action principle? “There is no justification for the particular choice of the 5-D curvature scalar as integrand of the action integral, from the standpoint of the restricted group of the cylindrical metric” (Pauli 1958 230).

All in all, the inconsistency between the peculiarity of the conditions imposed upon x^4 and the conservative **GR** on it are the signs of *ad-hocness*. A solution to this problem can be provided in the context of general skepticism towards extra dimensions and will be addressed in Part III, section 13.2. As a undesired consequence of Kaluza’s theory:

[32] *Electrons (and other charged microparticles) do not follow geodesics.*

Geodesics in 4-D are curves parameterized by λ that parallel transport their own tangent vectors: $\frac{dx^\mu}{d\lambda}$. The parallel transport is defined as: $\frac{D}{d\lambda} = \frac{dx^\mu}{d\lambda}\nabla_\mu$, so the equation of the geodesics is deduced from $\frac{D}{d\lambda}\frac{dx^\mu}{d\lambda} = 0$. The general form of the geodesic equation is:

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\sigma}\frac{x^\rho}{d\lambda}\frac{x^\sigma}{d\lambda} = 0 \qquad (86)$$

The first important test of Kaluza’s new unified theory was the analysis of geodesics in 5-D. In the vacuum theory, $T_{\mu\nu}$ encodes the kinematic energy of test particles. The ideal situation would be like this: a small, charged test particle in 5-D falls on a geodesic in

5-D and its projection in 4-D is the expected trajectory of a charged particle (typically *not* a geodesic). In order to estimate the geodesics of charged elementary particles, Kaluza faced a difficulty because his equation of geodesics is more complicated:

$$\frac{d^2x^m}{d\lambda^2}+\Gamma^m_{ab}\frac{dx^a}{d\lambda}\frac{dx^b}{d\lambda}=-\frac{p_4}{M}\kappa F^m_n\frac{dx^n}{d\lambda}-\frac{1}{2}\left(\frac{p_4}{M}\right)^2\phi^{-2}\nabla^m\phi \tag{87}$$

The last term creates an extra force due to the present of the uninterpreted field ϕ. Kaluza intended to infer the equation of a particle with mass M and charge q in curved spacetime in which an electric field tensor $F_{\mu\nu}$ is present:

$$\frac{d^2x^\rho}{dt^2}+\Gamma^\rho_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}=-\frac{q}{Mc}F^\rho_\mu\frac{dx^\mu}{dt} \tag{88}$$

In order to meet the requirement of a smooth transition from 4-D to 5-D, Kaluza wanted to keep only the first term in the RHS of (93) in order to identify (94) and (93) within the [29]. If the parameterization is $\lambda=\tau\cong t$ there are two ways to identify geodesics in 4-D with projections of geodesics from 5-D.

- a vanishing term $\nabla^\mu\phi$ or ϕ=constant;
- a small or vanishing term $\frac{p^4}{M}$, the "slow motion approximation" for massive particles.

Kaluza tried to avoid the former and opted for the latter. In this case the interpretation of (91) can raise difficulties, but also it constitutes a powerful tool for explaining **EM**. Two particles in 4-D having the same mass and the same initial conditions and differing only in respect of their charge will follow two trajectories that are both projections of a geodesic in 5-D. This is explained by the fact that u^4, their velocities on the fifth axis is different. Had we started with the small velocity approximation, we would want u^4 to be close to zero. The formalism applies only to relatively small velocities *and* to charges of $\rho/m \ll 1$, which seems kosher for all practical purposes. But this second approximation is unsatisfactory for atomic dimensions where u^4 is not at all small for a given density charge of electron or proton. In this case, the slow motion is no longer met and the motion of an electron is *not* a geodesic in 5-D as u^4 is enormously large. This means that Kaluza's theory would not work for subatomic particles.

If [29] does not hold, then electrons do not follow the geodesics in 5-D. So which path do they move on? It seems that the field alone cannot determine the paths of probe particles which is a disaster for a vacuum theory. Then, what does determine them? The only answer at hand is: the field ϕ. On one hand this is too arbitrary to be accepted and on the other hand it can be a source of further explanatory power. Maybe this scalar field hides something and can bring about new explanations or even predictions. We'll see that the problem of geodesics equation are fixed in Klein's theory where ϕ does not play any explanatory role. Kaluza never thought of setting this field to a constant as did Klein. And moreover he did not think of the **EM** vector potential A_μ as being the quotient $\frac{g_{\mu 4}}{g_{44}}$ as did Klein. These two ideas opened up a plethora of possibilities to Klein.

Chapter 11. Klein and quantum mechanics in five dimensions (1926)

11.1. Historical background

Although Kaluza's paper was known to the physics community (apart from Einstein, at least W. Pauli and N. Bohr had known of it), there were no attempts to pursue the idea further. His speculations on the quantum and statistical nature of the scalar field ϕ came too early and a major element was missing: *the wave description of quantum dynamics*. It has been said that Bohr suggested if a theory of space and time in four dimensions cannot describe quantum phenomena, maybe a theory of space and time in more than four dimension could.[154] In 1923-1925, Oskar Klein, a close collaborator of N. Bohr, working temporarily in Ann Arbor, Michigan, independently came upon the idea of the five-dimensional approach to unification in the context of the dynamics of charged particles in gravitational and electromagnetic fields.[155] Klein started from the aforementioned similarities between **GR** and **EM** after teaching a class on electrodynamics. In one of his lectures in 1969, Klein recollected the early 1920s:

> The similarity struck me between the ways the electromagnetic potentials and the Einstein gravitational potentials enter the [relativistic Hamilton-Jacobi equation for an electric particle], the electric charge in appropriate units appearing as the analogue to a [fifth] momentum component, the whole looking like a wave front equation in a space of [five] dimensions. This led me into a whirlpool of speculation, from which I did not detach myself for several years and which still has a certain attraction for me.
>
> […] I became immediately very eager... to find out whether the Maxwell equations for the electromagnetic field together with Einstein's gravita-

[154] This quote is Klein's recollection. See (Ekspong 1991 109).
[155] Klein's PhD thesis was on the statistics of the Brownian motion in strong electrolytes. In Ann Arbor, Klein was studying the anomalous Zeeman effect, the behavior of atoms in magnetic fields.

> tional equations would fit into a formalism of five-dimensional Riemann geometry (corresponding to four space dimensions plus time) like the four-dimensional formalism of Einstein. It did not take me a long time to prove this in the linear approximation, assuming a five-equation, according to which an electric particle describes a five-dimensional geodesic (Ekspong 1991 108, 109-110).

Lacking a proper training in relativity, Klein read Pauli's newly published *Relativitätstheorie* and realized that there is a way to infer Einstein's field equations from a five-dimensional theory, at least in a first approximation of "weak fields". As Pauli remembered, in those early years Klein did not intend to interpret the fifth axis in any way (Pauli 1958, 241). In 1925, he returned to Denmark and continued his collaboration with Niels Bohr. While in Copenhagen, Klein contacted a serious disease and was on forced bedrest in Stockholm between September 1925 and March 1926.[156]

While on leave from academic duties he realized that the component of momentum along the fifth dimension can be interpreted as being proportional to the electrical charge if the space is closed in the direction of the fifth dimension x^4 with a circumference of $.8\mathrm{x}10^{-30}$ cm, "far beyond the smallest distances observed" (Ekspong 1991 110). This assumes a new topology of the fifth direction. Physical quantities are periodic functions of the x^4 and measurable quantities are averages taken over the fifth axis and higher overtones correspond to states of high electric charges. In this respect the periodicity of x^4 was "the root of the quantal aspect of nature" (Ekspong 1991 110).

I think that had Klein's theory been known to Schrödinger, the whole history of quantum mechanics would have been different because Klein's result would encourage Schrödinger to pursuit the relativistic form of Schrödinger equation, at least. If Klein was

---

[156] Klein worked in Copenhagen and in Lund with N. Bohr between 1926 and 1930. In 1930 he took a position at Stockholm University where he remained till his retirement in 1962.

right, then any quantum effects in electromagnetism can be explained as the behavior of the de Broglie "matter" wave in extra dimension. It was not clear whether the dynamics of quantum particles in general could be explained by the dynamics of fields in 5-D.

This may sound like Kaluza's conclusion at the end of his 1921 paper and it would have been music to the ears of Einstein as well. The history is quite different. Kaluza never published in physics after 1921; according to several accounts, Klein learned about Kaluza after he wrote a first version of the paper (sometimes in the spring of 1926); Einstein did not like Klein's ideas and never quoted Klein's 1926 paper, except in an incidental footnote in the 1938 paper co-written with P. Bergmann. Klein learned about Kaluza only in 1926 from W. Pauli, and not from Bohr, long after he had finished writing a first draft of his own paper:

> When Pauli came to Copenhagen […], I showed him the manuscript on five-dimensional theory and after reading it he told me that Kaluza some years before had published a similar idea in a paper I had missed. So I looked it up [...] but I read it carelessly but quoted [it] in the paper I then wrote in a spirit of resignation. […] In the paper I tried, however, to rescue what I could from the shipwreck and in the same time to learn as much as possible from Schrödinger and de Broglie. Klein recounting the 1920s in (Ekspong 1994 111).

Notwithstanding this personal dissatisfaction, Klein's paper (including the acknowledgements to Kaluza and to de Broglie) was published in April 1926 in *Zeitschrift für Physik* and a note appeared in October in *Nature*.[157]

I found that the literature on Kaluza and Klein, written mainly by enthusiast physicists, is confusing and these episodes both cry for a deeper philosophical investigation. Klein's 1926 paper is presented as a continuation of Kaluza's ideas as seemingly his pub-

[157] See (Klein 1926, 895-906) and translations in (Appelquist, Chodos, and Freund 1987; Sabbata and Schmutzer 1982; O'Raifeartaigh 1997). The short paper in *Nature* is (Klein 1926, 516).

lished paper was inspired by Kaluza's work. At a first sight, Kaluza's and Klein's theories are strongly related and many authors happily admit that Klein merely extended Kaluza's theory.[158] This is false, as I argue through this chapter and in Part III. What is more serious is that in fact almost all exposés of Kaluza and Klein are not usable because many confuse Kaluza with Klein—and vice versa. Although Kaluza's theory has had a Lagrangian formulation, Kaluza never wrote the action for it. Although one can infer from Kaluza the invariance to coordinate transformation, Kaluza never talked about it. Many authors talk of "Kaluza-Klein" as if the theory was originally formulated as such. Here is an example of another patent confusion. Appelquist, Chodos and Freud, the three editors of an excellent collection of papers on Kaluza and Klein theories, committed even a more serious blunder by talking about Kaluza's "action", Kaluza's "compactification" of x^4 on a circle S^1, the Fourier expansion and of the condition ϕ=1!!![159] I showed that there was no compactification in Kaluza, so no Fourier expansion. Even in Klein, compactification as such appears as a conclusion of a longer argument in the 1926 paper and as a supposition in the note to *Nature*. Fourier expansion is not present in Klein either, and it is explicitly written down only in Einstein and Bergmann's paper of 1938 and discussed again in the context of Yang-Mills theories in the late 1950s. None of these features are present in Kaluza's paper and many of the claims are blatantly false. Presumably because these papers are not papers in the history of science they conflate the historical context with the conceptual context.

An exception is a report by H. Goenner and D. Wünsch written at the Max Planck Institute in which the authors carefully made a distinction between Kaluza's and Klein's

---

[158] See (Kragh 1984 410). Some authors in a more superficial manner confuse Kaluza and Klein: see (Blagojevic 2002 295 eq 10.2b is not written by Kaluza). The confusion of action is present also in (Duff 1995 4; O'Raifeartaigh 1997 48). Susskind also talks about Kaluza and in fact it refers to Klein.

[159] (Appelquist, Chodos, and Freund 1987 4-7) On p. 6 another blunder: "Kaluza arbitrarily set $\phi^{(0)}$=constant.

contributions.[160] In my work I emphasize other differences between Kaluza and Klein and focus especially on the different forms of scientific unification their theories illustrate. My presentation sheds some light on the historical context of Kaluza and Klein's work. This is why I describe details of Klein's approach and avoid perfunctory conclusions, given the fact that the difference between Kaluza and Klein play a crucial part in my argument. In Klein, there are significant improvements both in the formalism itself and in the conceptual scheme, especially in the second part the 1926 article, and in the note to *Nature*. However, by using his new hypothesis about the fifth direction, Klein radically differs from Kaluza. Several aspects that constitute takeoffs from Kaluza's formalism will be discussed in Part III.

Action principle in 5-D. We saw that in 1916, Hilbert applied the action principle to infer the field equations and within weeks Einstein himself started to heavily use the action principle. From the beginning, the action principle played a major heuristic role in Klein and it reflected his way of drawing the parallel between the electromagnetic and the gravitational fields. Even before reading Kaluza's paper, Klein's incipient idea was to associate to the momentum on the fifth axis a quantity similar to the mass.[161] Klein started from the action of an electron of charge e and mass m in a combined electromagnetic and gravitational field:

---

[160] Goenner's and Wünsch's material (Wunsch and Goenner 2005) was presented at the Tenth Marcel Grossmann Meeting and it was published in the *Proceedings of the MG10 Meeting held at Brazilian Center for Research in Physics (CBPF),* Rio de Janeiro, Brazil, 20-26 July 2003. I did not have access to this resource, but I used here a preprint with the logo of the Max Planck Institute. See also note ###

[161] I have had no access to Klein's archives. I rely here on a presentation by Lars Brink at the conference *Oskar Klein Meeting: D > 4, t = 75,* (Ann Arbor, October-November 1998), held 75 years after the time when Klein worked on the fifth dimension. See the website at: http://feynman.physics.lsa.umich.edu/klein/newklein.html. I prefer to discuss the 1926 paper instead of Klein's previous notes.

$$S = \int d\tau \left[\frac{m}{2} g_{\mu\nu} \dot{x}^{\mu} \dot{x}^{\nu} + eA_{\mu} \dot{x}^{\nu} \right] \tag{89}$$

and speculated that A_μ contains a hidden index such that $A_\mu = \kappa g_{\mu 5}$ and that e is the aspect of a velocity on the fifth axis such that $\frac{e}{\kappa m} = \dot{x}^5$. By this, the action has a more homogeneous form:

$$S = \int d\tau \left[\frac{m}{2} g_{\mu\nu} \dot{x}^{\mu} \dot{x}^{\nu} + m g_{\mu 5} \dot{x}^{\mu} \dot{x}^{5} \right] \tag{90}$$

which suggests a 5-D action of the g_{mn} field. Klein was baffled by the term g_{55} and its absence from the 5-D action as well as by the interpretation of the suggested: $p^5 = m(\dot{x}^5 + g_{5\mu}\dot{x}^{\mu}) = \frac{e}{\kappa} + m g_{5\mu}\dot{x}^{\mu}$.

In the published paper, Klein has been more rigorous than Kaluza in justifying the metric $g_{mn}^{(5)}$ in 5-D. Instead of guessing an expression for the Ricci tensor, he explicitly employed a 5-D action of the scalar field $R^{(5)}\sqrt{-|g|^{(5)}}$ similar to what Hilbert did. The Lagrange density of fields depends only on $g^{(5)}$ and its first order derivative $\partial_i g^{mn}$:

$$S_{Klein} = \int R^{(5)} \sqrt{-|g^{(5)}|}\, d^5 x \tag{91}$$

where $R^{(5)}$ is a Ricci-like invariant scalar (similarities with **(36)** are obvious) defined by (Klein 1926 897):

$$R^{(5)} = \hat{g}^{mn} R_{mn} = \hat{g}^{ik} \left(\partial_k \Gamma^{i\mu}_{\mu} - \partial_{\mu} \Gamma^{ik}_{\mu} + \Gamma^{i\mu}_{\nu} \Gamma^{k\mu}_{\mu} - \Gamma^{ik}_{\mu} \Gamma^{\mu\nu}_{\nu} \right) \tag{92}$$

where the assumptions needed for the derivation of the action in 5-D are identical to those used by Hilbert in 4-D. From this perspective, the transition to 5-D is as conservative as possible.

Coordinate transformation. At this point Klein realized that a departure from Kaluza is necessary and some extra assumptions are needed to simplify the variation of the 5-D action. In the language used in the 1920s, Klein looked for invariance in the coordinate transformations. In a more modern language, symmetries need to be imposed on the 5-D manifold. He supposed that the four original spacetime directions do not transform in the same manner as x^4 does. The same can be said about the stationary solutions in **GR** where time is less "spatio-temporal" than the other spatial directions. Klein imposed two weaker conditions on the **5-D** coordinate system:

[33] *The first four coordinates are identical to the ordinary spacetime coordinates;*

[34] *The field g does not depend on the fifth coordinate.*

It is worth mentioning that in the second part of the 1926 paper and in the note to *Nature*, Klein would replace [34], very similar to Kaluza's **(CYL),** with the stronger condition of compactification (**COMP**), i.e. rejection of the linear structure of the fifth axis (x^4). One can see that Klein's (**COMP**) is far less intuitive and stronger than Kaluza's (**CYL)**, but it dramatically changes the theory's unificatory power. A possible motivation for assuming (**COMP**) was the fact that it was not observable.

Before entering into the details of (**COMP**), a transgression in the topic of coordinate transformation is in order here.[162] As in Kaluza, the condition [34] can be easily formulated as $\partial_4 g_{mn}^{(5)} = 0$. As I mentioned in analyzing Kaluza, this is not generally covariant, although it holds for a large class of coordinate systems. How do we convert [33] to the language of the action principles? The answer is simple. We can think of a special class of coordinate transformations that leaves the 5-D action invariant whenever this extra symmetry of g is added.[163]

I mentioned already that there was something special about x^4 in Kaluza that frustrated Einstein and other physicists. I suggest that Klein's argument based on compactification is similar to the analysis of ordinary **GR** with static solutions. The fifth axis is special and its special regime can be intuited by an analogy with the special character of the time coordinate in **SR** and especially in **GR**. Once again an analogy with GR would have helped Klein.

*Space and time transformations in **SR***. It is usually said that spacetime is space and time unified under one concept in Minkowski. Indeed, one of the popularized stories is that after **SR**, space and time have played the same role in physics. This is only partially true at best and there are enough aspects of time within **SR** that would not fit this simplified image. Is time special in **SR**?[164] I pick a specific answer to this question. Because of the indefinite character of the metric and because of its signature, the inhomogeneous Lorentz transformations:

[162] Lacking any historical evidence, I can only speculate that **(COMP)** had been part of Klein's theory even before he learned of Kaluza's paper, and that [34] was added in the published paper in 1926.
[163] Klein did not talk about symmetries, but about invariance of g which can be converted into a talk about symmetries.
[164] For a discussion on why time is special, see (Callender 2008).

$$\overline{x}^{\mu} = x^{\nu}\mathbf{\Lambda}^{\mu}{}_{\nu} + \mathbf{C}^{\mu} \tag{93}$$

two events can be separated by a zero distance even if they are not identical. This means a translation in spacetime is not the same as a translation in space or as a translation in time. Take for example a transformation as simple as rotation. Lorentz transformations does not permit all kinds of rotation and translation invoking the *t* axis, while a pure spatial metric permits all continuous rotations invoking spatial directions (i.e. they are part of the Lorentz group).

*Time and static solutions in **GR**.* How **GR** handles time is an interesting philosophical discussion, especially in the context of the Hamiltonian formulation. I focus on something different, the so-called "stationary" solutions, paramount to early cosmological models. I believe that they illustrate the importance of time in **GR** and that they were useful in the genesis of Klein's theory. Stationary universes were known to Einstein and discussed by H. Kramers who was quoted in Klein's paper.[165] A stationary spacetime means that there exists a coordinate system where the metric tensor is independent of x^0, i.e. $g_{0a}=0$ for a=1,2,3. It is usually said that the stationary solutions are possible where a group of isometries have orbits with timelike curves. The group of isometries expresses the "time translation symmetry" of spacetime (Wald 1984 119). In more technical terms, a metric is stationary if it has a timelike Killing vector field. If the Killing vector field is orthogonal to a family of spacelike surfaces of constant time, it is also considered static.

A stationary spacetime permits only a special class of transformations:

[165] (Kramers 1922). In fact it is now known that very few solutions are stationary, comparatively speaking. Our own universe is not stationary, but this is a later finding. Real stationary solutions include the Ernst vacuum, some dust solutions (van Stockum), disks (Meinel-Neugebauer), the Bonor beam etc.

$$\begin{cases} \overline{x}^A \to x^A + \Xi^A(x^B) \\ \overline{x}^0 \to x^0 + \Lambda(x^A) \end{cases} \tag{94}$$

where Ξ are differentiable, but arbitrary functions of space-like coordinates only and Λ is a arbitrary function of spatial coordinates only that has a line element without terms in $dt\, dx^A$:

$$ds^2 = h_{AB}(x^C)dx^A dx^B - V^2(x^C)\left(dx^0\right)^2 \tag{95}$$

where capital Latin indices run from 1 to 3 and C is a dummy index (the scalar field V is also a function of Ξ and Λ). Both h and V do not depend on time. It is easy to see that in this case vectors split in temporal and spatial components and that the symmetry groups of the space directions and of those of time directions differ.

11.2. Klein's metric

Although Klein could have been inspired by the "stationary spacetime", his argument is based on the action principle, too.

If ξ are infinitesimal displacements of coordinates such that $\delta x_\mu = \xi_\mu$, then the variation of the metric is:[166]

$$\delta g_{mn} = \partial_m \xi^r g_{rn} + \partial_n \xi^r g_{mr} + \xi^r \partial_r g_{mn} \tag{96}$$

The variations of the metric tensor splits into two sectors:

$$\begin{cases} \delta g_{\mu\nu} = \partial_\rho g_{\mu\nu} \xi^\rho \\ \delta g_{4\mu} = g_{44} \partial_\mu \xi^4 + \partial_4 g_{4\mu} \xi^4 \end{cases} \tag{97}$$

Given [33] and [34] and given that the displacements are infinitesimal, one can calculate:[167]

[166] See **(47)**.

$$\begin{cases} \delta g_{44} = 0 \\ \delta g_{\mu 4} = g_{44}\partial_{\mu}\xi \end{cases} \tag{98}$$

Results from the coordinate transformation analysis. Immediate consequences of these relations (Klein 1926 896) follow:

- The g_{44} is a constant:

$$g_{44} = \phi = \text{constant} \tag{99}$$

- The transformations of coordinates are:

$$\begin{cases} \bar{x}^4 = \alpha x^4 + \xi(x^{\mu}) \\ \bar{x}^{\mu} = \bar{x}^{\mu}(x^{\mu}) \end{cases} \tag{100}$$

- Given how the metric changes, the invariants of (106) are the normal diffeomorphisms of 4-D transformation:

$$\overline{g}_{\mu\nu} \rightarrow g_{\mu\nu} - g_{\nu 4}\partial_{\mu}\xi - g_{\mu 4}\partial_{\nu}\xi - \partial_{\mu}\xi\partial_{\nu}\xi \tag{101}$$

  And the invariant of the x^4 is a gauge invariance that splits into two terms characterized by α and ξ.

- The 5-D Ricci tensor is written as a sum of a 4D Ricci quantity and another tensor proportional to $F^{\mu\nu}F_{\mu\nu}$. After some calculations, one obtains a Ricci scaler (Blagojevic 2002 296):

$$R^{(5)} = R^{(4)} + \frac{g_{44}}{4}F^{\mu\nu}F_{\mu\nu}\,. \tag{102}$$

- The quotient vector $\mathbf{B}_{\mu}=g_{\mu 4}/g_{44}= g_{\mu 0}/\phi$ transforms like:[168]

---

[167] Some other assumptions are necessary to perform calculations here. See (Blagojevic 2002 295; O'Raifeartaigh 1997 51). Although both authors refer to Kaluza, in fact they analyzed Klein's assumptions.
[168] In projective geometry the vector **B** is written as a 5-D vector. See (Veblen and Hoffmann 1930#; Bergmann 1942 265).

$$\overline{\mathbf{B}}_\mu = \frac{\overline{g}_{\mu 4}}{\overline{g}_{44}} \to \frac{g_{\mu 4}}{g_{44}} + \partial_\mu \xi \tag{103}$$

At least at the level of the transformation group, this looks similar to the transformations of h_{ab} in (101). For such a transformation, the metric tensor has the form:

$$g^{(5)}_{mn} = \begin{pmatrix} g_{\mu\nu} + \mathbf{B}_\mu \mathbf{B}_\nu & \mathbf{B}_\mu \\ \mathbf{B}_\nu & 1 \end{pmatrix} \tag{104}$$

and looks like the covariant form of a field. Here the quotient vector **B** helps to set the g_{44}=1 and to establish the signature of the metric to (– + + + +), i.e. the fifth dimension is spatial, so did Kaluza.[169]

Like many other components of tensors, the field **B** would have no physical meaning in itself. Klein proceeds here to the first identification. As components of **B** transforms like the covariant component of the electromagnetic tensor $F_{\mu\nu}$, the simplest way is to set the four components of **B** proportional to the **EM** vector potential A_μ:

$$B_\mu = \frac{g_{\mu 4}}{g_{44}} \beta A_\mu \tag{105}$$

where β is a constant of proportionality.

Thus, the form of $g^{(5)}$ is:

$$g^{(5)}_{mn} = \begin{pmatrix} g_{\mu\nu} + \beta^2 A_\mu A_\nu & \beta A_\mu \\ \beta A_\nu & 1 \end{pmatrix} \tag{106}$$

Klein's metric has several new features compared to Kaluza's. First, the metric of the 4-D spacetime has an electromagnetic component $\beta^2 A_\mu A_\nu$ (more on this in Section 14.3). Second, Klein realized that the presence of the new, un-interpreted constants in the

[169] In Klein, the GR has the signature (- + + +), so for the opposite convention, one has to choose g_{44}=-1.

metric is not a good practice and that he still had a “free” constant β. Both Kaluza’s and Klein’s theories generated surplus structure as constants and structures that are serious hindrances if one wants to interpret the theory literally. Setting β to a definitive value is simpler and Klein dealt with it. The field $g_{44}=\phi$ constitutes a different story. It was a problem in Kaluza; Klein set it to a constant, typically 1; later on it played an important role in Thiery’s theory of a variable gravitational constant. If one keeps the scalar in the form (110) the metric is:

$$g_{mn}^{(5)} = \begin{pmatrix} g_{\mu\nu} + \phi\beta^2 A_\mu A_\nu & \phi\beta A_\mu \\ \phi\beta A_\nu & \phi \end{pmatrix} \tag{107}$$

An alternative notation, used in the “projective geometry” formulation of Veblen and Hoffmann in which the metric is transformed by $x^4 \to e^{x^4}$ (Duff 1995, 22-35):

$$\hat{g}_{mn}^{(5)} = e^{\varphi/\sqrt{3}} \begin{pmatrix} g_{\mu\nu} + e^{-\sqrt{3}\varphi} A_\mu A_\nu & e^{-\sqrt{3}\varphi} A_\mu \\ e^{-\sqrt{3}\varphi} A_\mu & e^{-\sqrt{3}\varphi} \end{pmatrix} \tag{108}$$

The equation of the field ϕ, present in Kaluza and dismissed by Klein, will come back later in another context. Setting $g_{44}=\phi$ to a constant seemed arbitrary, but it was used for computational reasons. Kaluza’s theory suffered from being too general, but by hypothesizing ϕ constant it helped Klein to improve Kaluza’s theory. Kaluza left this field uninterpreted, but Klein chose to set it to a constant in order to solve the problem of geodesics.

Probes in 5-D and the geodesics. The first good news for Klein was that the metric yielded the right form of the geodesics in 5-D. Indeed, Klein added to the action a Lagrange density for the motion of *n* free charged particles. The total Lagrange density in the presence of fields and *N* probe particles is (Klein 1926 899):

$$\mathcal{L} = \mathcal{L}_1 + \sqrt{-g^{(5)}}\,\kappa \sum_{i=1}^{N} g^{mn} \frac{dx_i^m}{d\lambda} \frac{dx_i^n}{d\lambda} \tag{109}$$

where $\mathcal{L}_1$ is the Lagrangian of the field, κ is a constant of proportionality and the last term corresponds to the kinetic energy of the probe particles. Similar to Kaluza's identification of u^4 with the charge **(91)**, in order to derive the geodesics in 5-D, Klein interpreted the velocity on the fifth axis as proportional to the charge of the particle:

$$\begin{gathered} u^4 = \pm \frac{e}{c} \frac{1}{\dfrac{d\tau}{d\lambda}} \\ \frac{d\tau}{d\lambda} = \sqrt{M} \text{ or } \sqrt{m} \end{gathered} \tag{110}$$

where as usual: $d\tau = \frac{1}{c}\sqrt{-ds^2}$ is the proper time in 5-D, λ is a parameter of the geodesics and *e* is the electrical charge of the electron and M is the mass of the hydrogen "nucleus" and m the mass of electron.[170] Klein thought that the charged particles on geodesics are the hydrogen nucleus and electrons. So this is why he set two signs, - for electron, + for the "hydrogen nucleus". From the Ricci tensor, Klein inferred the 5-D geodesics. On such geodesics, the Lagrange function $\mathfrak{L} = \frac{1}{2}\left(\frac{ds}{d\tau}\right)^2$ provides the definition of the 5-D momentum:

$$p_i = \frac{\partial \mathcal{L}}{\partial \left(\dfrac{dx^i}{d\lambda} \right)} \tag{111}$$

---

[170] He was thinking of the proton, a positively charged particle without any neutron.

As there is no explicit dependence of $\mathfrak{L}$ on x^4, we will always have a constant momentum on the fifth axis. The calculations render for an electron:

$$p_4 = \pm \frac{e\sqrt{\phi}}{c\sqrt{2\kappa}} \tag{112}$$

As ϕ is kept constant in spacetime, p_4 has the same value at any point of spacetime.

$$p_4 = \beta\left(\pm\frac{e}{c}\right) \tag{113}$$

Up to now the constant β was not determined. From (118) Klein arbitrarily set the conditions of masses and charges and for the field $\phi = \frac{c\sqrt{2\kappa M}}{e}$ such that p_4 is normalized to 1:

$$p_4 = \begin{cases} -1 \text{ for electrons} \\ +1 \text{ for H-nucleus} \end{cases} \tag{114}$$

In a zeroth interpretation, the electron has always a constant motion on x^4 with a given heliticity and the hydrogen nucleus, i.e. the proton has a constant motion, too, with the same velocity but with the opposite heliticity. Klein's suggestion is that charged particles move in a very distinct way along x^4: with the same constant velocity and in the two possible directions. Klein used an analogy: think of a particle moving in on a circle in 3-D with a constant speed; project this motion to an arbitrary plane and you will get the illusion of different, even accelerated motions. Again, projection in 4-D of a very simple 5-D dynamics could be very complicated (Kragh 1984 1026). Moreover, the field ϕ is always constant on the x^4 direction.

11.3. Klein's second argument: compactification and waves in 5-D

De Broglie's matter wave. The second major step is the usage of a wave front equation in five dimensions which explains the quantization of charge. Klein studied the differential form of a "ray" of a wave function Ψ that obeys the wave equation in 5-D:

$$\Box_g \Psi = 0 \tag{115}$$

(where $\Box_g = g^{mn}\left(\frac{\partial^2}{\partial x^m \partial x^n} - \Gamma^k_{mn}\frac{\partial}{\partial x^k}\right)$ is the wave operator in 5D, i.e. the d'Alembertian in its covariant form[171]) that is covariant, i.e. it does not depend on the system of coordinates. Equation (121) cannot be solved in general, even under the restrictions (105) and (106). In the so-called "geometrical optics approximation" where the variation of the field Ψ is much smaller than that of the curvature, the wave solution has a simple expression of an exponential depending on a phase factor only:

$$\Psi = A\exp\left[i\omega\Phi(x^\mu)\right] \tag{116}$$

with an amplitude A with small derivatives. In the case of geometrical optics, the differentiation of (121) produces a quadratic term and a linear term in ω, without a free term. The quadratic term in ω^2 is:

$$g^{mn}(\partial_m\Phi)(\partial_n\Phi) \tag{117}$$

and the linear term in ω is:

$$g^{mn}(\partial_m\partial_n\Phi - \Gamma^{mn}_r\partial^r\Phi) \tag{118}$$

---

[171] A more compact form of the wave massless equation is $\frac{1}{\sqrt{-|\hat{g}|}}\partial_m(\sqrt{-|\hat{g}|}\hat{g}^{mn}\partial_n)\Psi = 0$. Klein used a more general form in which a^{mn} are some functions of the coordinates, but he later implicitly identifies a with g.

The central point of the wave-particle analogy of de Broglie is the definition of the momentum by the operator "nabla" $\hat{p} = -i\hbar\nabla$ (Dongen 2002 5):

$$\hat{P}_m = \frac{\partial}{\partial x^m} \tag{119}$$

that acts on a wave function Ψ. So rays are null geodesics of the differential form: $g_{mn} dx^m dx^n = 0$.

Similar to Kaluza, Klein was forced to discuss elementary particles in two possible cases. From (119), one can see that the only factor dependent on x^4 in the wave equation is the phase Φ. Hence:

$$\Phi = \pm x^4 + S(x^0, x^1, x^2, x^3) \tag{120}$$

Where S is a function of the first four coordinates. From (126) the wave function can be furthermore separated:[172]

$$\Psi = \exp\left[\pm i\omega x^4\right]\psi(x^\mu) \tag{121}$$

CASE 1: **Large ω.** For ω large enough, the wave operator will have only the quadratic term such after differentiation the remainder is an equation of the phase Φ is (Klein 1926 900-902):[173]

$$g^{mn}\left(\frac{\partial\Phi}{\partial x^m}\frac{\partial\Phi}{\partial x^n}\right) = 0 \tag{122}$$

As the momentum in the Lagrangian formulation is defined as:

$$p_i = \frac{\partial \mathcal{L}_\Psi}{\partial \frac{dx^i}{d\lambda}} \tag{123}$$

[172] Here c is the speed of light in vacuum.
[173] For details of this approximation in the case of the **EM** wave, see (Wald 1984 71).

one can treat the wave function as a scalar function that satisfies (121), use a translation to a Hamiltonian formulation of **GR** of a scalar field U and then infer the equation for rays. After some calculations, Klein inferred a very interesting property of the Hamiltonian in 5-D: the total Hamiltonian $\mathfrak{H}_\Psi$ of this form is zero because of (129).

$$\mathfrak{H}_\Psi = 0 \tag{124}$$

As the definition of the Hamiltonian is:

$$\dot{p}_m = -\frac{\partial \mathfrak{H}_\Psi}{\partial x^m}; \dot{x}^m = \frac{\partial \mathfrak{H}_\Psi}{\partial p^m} \tag{125}$$

the ray equation becomes:

$$\frac{dp_4}{d\lambda} = 0; \frac{dp_\mu}{d\lambda} = \frac{1}{2}\partial_\mu g_{mn}\frac{dx^m}{d\lambda}\frac{dx^n}{d\lambda} + p_0\frac{d\phi_m}{d\lambda}\frac{dx^m}{d\lambda} \tag{126}$$

One can write the Lagrangian associated to this Hamiltonian.

On the other hand, the equation of motion of a charged particles is: $\mathcal{L} = \frac{1}{2}\left(\frac{d\theta}{d\lambda}\right)^2 + \left(\frac{ds}{d\lambda}\right)^2$. If rays coincide with the particle’s trajectory, then the Lagrangian of the wave has the same form as the Lagrangian for the particle. Klein used the conservation of phase Φ along a closed trajectory in the fifth dimension:

$$\omega \oint p_4 dx^4 = 2\pi n \tag{127}$$

Where n is any natural number. As the Hamiltonian of this wave is zero, the phase is conserved. This gives the condition of stationary of phase.

CASE 2 (the Klein-Gordon equation) Small ω. For this case, it is easy to see that second term in (124) can be neglected. Klein inferred the Schrödinger equation and, for the first time, the form of the relativistic wavefunction for a spinless particle, later named "the Klein-Gordon equation" (Klein 1926 901-902).

For a static electrical potential V depending on space only and for a metric that is slightly perturbed from its Minkowski value, one can suppose that (127) can be further separated:

$$\Psi = \exp\left[-2\pi i(\frac{x^4}{h} - \nu t)\right]\psi(x^\mu) \tag{128}$$

where ω=2π/h. This equation leads to:

$$\left[\triangle + \frac{4\pi^2}{c^2h^2}\left[\left(h\nu - eV\right)^2 - m^2c^2\right]\right]\psi = 0 \tag{129}$$

Klein thought that his equation applies to electrons, but this is blatantly false. The electron is not a spinless particle.[174] Klein endeavored to connect quantum results with the analysis of geodesics in 5-D. After performing the differentiation with respect to x^4, the wave-equation takes the familiar Klein-Gordon form: $\nabla_\mu \nabla^\mu \Psi - m^2\Psi = 0$ with $m = \frac{e}{\sqrt{2\kappa c^2}}$ which is a quantum result in 4-D. The field A_μ is incorporated in the covariant derivative in the usual way.[175] This equation has a manifestly covariant form and suggests that the dynamics of a wavefunction in 5-D is projected in 4-D as the dynamics of a relativistic scalar field.

[174] This equation was published in the same year by Klein, V. Fock and Gordon (allegedly, Schrödinger had first discovered and immediately rejected it in 1925 because it could not explain spin). Klein's manuscript was submitted to the editors of *Zeitschrift für Physik* in April 1926, whereas Fock's and Gordon's in July, and in September. Fock used a 5-D formalism, very similar to Klein's. In fact, the classical scalar field used by Klein here does not exist in nature, although it is related to the quantum field of pions. The Dirac equation that describes a spin ½ particle (although not the electron!) was discovered in 1927.

[175] I do not want to insist on the derivation of the Klein-Gordon equation, as it is described and discussed in almost all textbooks. Not much attention has been paid to the fact that the Klein-Gordon equation originated in the 5-D formalism. For an excellent historical analysis, see (Kragh 1984, 1024).

In less than half of a page Klein was able to infer the Klein-Gordon equation and to express the Planck constant as a function of other constants of nature. It is amazing that Klein did not take this result seriously as he was mesmerized by the five-dimensional theory itself. Interestingly enough, in less than an year several physicists have showed that the Klein-Gordon equation can be inferred without any assumptions about the fifth dimensions.[176]

11.4. The new argument with COMP

I now go back to the main result of Klein discussed in **CASE 1**. In the note to *Nature*, instead of postulating the same values for the physical fields on x^4 as Kaluza had, Klein took a different stance: based on the stationarity of the wavefunction, he supposed that the axis is curled with a very small radius. In *Nature*, Klein explicitly discussed this condition: "The charge q, so far as our knowledge goes, is always a multiple of the electronic charge e, so that we may write $p_4 = n\frac{e}{k}$ with $n \in \mathbb{Z}$. This formula suggests that the atomicity of electricity may be interpreted as a quantum theory law." (Klein 1926, 516).

He hinted at the idea that the momentum along x^4 is always quantized and this reverse the direction of the explanation. From (133) and (119) one can infer:

$$p_4 = ne / c\sqrt{2\kappa} = n\hbar / \lambda_4 \tag{130}$$

where λ_4 is the radius of the closed circle on x^4. If one knows the quanta of electrical charge, from (136) one can deduce the compactification factor λ_4=.8 $\cdot 10^{-30}$ cm. Klein identified geometrically the points P and P' separated by $2\pi\lambda_4$ and rejected the linear

---

[176] L. Landau, G Gamow and A. Iwanenko had important contributions. The fact that the Klein-Gordon equation could be derived in other ways had a negative impact on the credibility of the Kaluza-Klein theory. See details in (Kragh 1984).

geometry of x^4. An immediate consequence is that one can think of a circle on x^4 (explicitly stated in the note to *Nature*).

Topology. Compactification is a major change in the topology of $\mathcal{M}$. Klein took seriously the non-Euclidian topology of the x^4. The compactification of the fifth dimension explains why it is not visible and I take it as being less ad-hoc than Kaluza's **CYL**. The topology of the five-dimensional space is no longer the topology of $\mathbb{R}$. Instead of postulating the physical fact that no fields depend on x^4, Klein simply identified points P and P'. As quantum formalism suggests in the case of the hydrogen atom, a stationary wave function could provide the condition of quantization. Points P and P' are *identical* in this topology if their fifth coordinates differ by a multiple of a fixed quantity.

$$\textbf{COMP:}\ x'^4 = x^4 + 2\pi\lambda_4 \qquad (131)$$

As **(COMP)** is not a coordinate variant of the theory, the new structure of x^4 is not a mere alternative *representation*, but it reflects the structure of the reality (unlike, for example, the case of polar coordinates where there are no transformations that remove the symmetry $S(1)$ and linearize x^4). The new manifold is not invariant anymore under the group GL(5), rather under the group $GL(4)\otimes S_1$ where S_1 is the group of a translation $x_{4'} = x_4 + \psi(x^i)$. If two particles have the same initial condition in 4-D (same positions and same velocities as components of the x_0^μ) but rather different ratios q/M, they will follow geodesics in 5-D. This is an improvement over Kaluza's approach. Klein's metric does not need the small velocity approximation used by Kaluza and solves the problem of geodesics.

The following are Klein's arguments in the two 1926 articles. I discuss these topics related to explanation in Part III.

The first argument in the 1926 article

[35] *The first four coordinates are identical to the ordinary spacetime coordinates; [33]*

[36] *The field g_{mn} does not depend on x^4 , i.e. [34]*

Result

[37] *The EFE equation and the Maxwell equations.*

The argument of the second part in the 1926 paper, part II.

[38] *The de Broglie hypothesis: the behavior of a particle is described by its "matter wave" function*

[39] *The "matter wave" function on x^4 is stationary.*

[40] *The wavelength is very small, i.e. the frequency of the wave function is large.*

Results

[37] as before

[41] *The fifth dimension is compactified **(COMP)***

[42] *Electrons move on geodesics*

The Argument in *Nature*

[37] as before

[43] *The fifth dimension is compactified **(COMP)**, i.e. the symmetry of the x^4 axis is S_1.*

Results

[37] as before

[44] *The quantization of electrical charge.*

[45] *Schrödinger equation and Klein-Gordon equation*

[46] *The symmetry of the **EM** theory U(1) is a consequence of the symmetry of spacetime manifold* $\mathbb{R}^4 \otimes S_1$.

The consequence of the initial argument was promoted to a hypothesis of the new argument and the hypothesis of the old argument (the quantization of charge) became a consequence and *explanandum* of the new one. The new hypothesis **(COMP)** is then used to explain the quantization of charge and the new symmetry group of the theory. The smallness of λ_4, being less than the Planck length, explains also why extensions on x^4 cannot be observed by macroscopic observers. In the new argument, given the value of λ_4 (**COMP)** explains **(CYL)**, the quantization of charge and the lack of observable effects at macroscopic scale. Without the compactification of the fifth dimension there is no explanation for the electrical quantization, a major topic to be discussed in Part III.

PART III. THE KALUZA-KLEIN UNIFICATION AND EXPLANATION

I discuss unification and explanation in Kaluza and Klein theories by addressing a general question:

[47] *How do the unificatory and explanatory mechanisms work together?*

The interplay between unification and explanation is easier to understand if we think both of these two theories as a two-stage process well illustrated by Kaluza and Klein.

The first is the "unificatory stage": here Kaluza and Klein add structure to the 4-D spacetime structure in order to achieve unification. How do we unify two physical interactions? In Chapter 12 I discuss the specifics of the unification of electromagnetic and gravitational interactions by: (a) contrasting it to reduction and (b) by ranking known cases of unification. The first questions addressed in this chapter are:

[48] *Is reduction possible between **EM** and **GR**? If not, why?*

I argue that the strategy of reducing **EM** interaction to gravitation or the other way around does not work and this makes room to unification. Second, I discuss T. Maudlin's ranking of unification and ranking Kaluza and Klein in this scheme. In the context of Mauldin's ranking of unification, I address these questions:

[49] *In what sense are Kaluza's and Klein's theories unificatory?*

[50] *What is specific to Kaluza's and Klein's unifications in comparison with other popular attempts and what do they teach us about unification in general?*

In the following two chapters I develop arguments pertaining to the importance of unification in Kaluza and in Klein. In Chapter 13 I argue that Kaluza achieves unification by showing in what respect it is not trivial. In Chapter 14 I show why Klein obtains a higher degree of unification compared to Kaluza. I reveal the improvement that Klein brought in respect of the unificatory power of the theory. Some general questions to be addressed are related to the novelty and specificity of both attempts:

[51] *In what sense is Klein's unification better than Kaluza's?*

[52] *What is specific to Klein and is not present in Kaluza?*

In respect of explanation, a general question is how much explanation one can acquire from unificatory theories. By taking a superficial look at scientific theories in general, we saw that some explain too much and some do not explain enough. Spacetime theories were frequently under scrutiny with respect of their explanatory power. The general question here is:

[53] *Are spacetime structures explanatory?*

I show in what sense spacetime theories are explanatory without involving causation. In the following two chapters I discuss explanation in Kaluza and in Klein—related to the enrichment of the spacetime structure. I discuss the way in which this new structure of spacetime is endowed with explanatory powers. The question of the last two chapters is:

[54] *What is unified and what is explained in Kaluza's and in Klein's theories?*

I argue that in Kaluza and in Klein spacetime structures play an explanatory role. In other words I argue for a form of non-causal explanation. In Chapter 16 I propose a method of analyzing the power of explanation both in Kaluza and Klein as intended consequences of the unificatory stage. In the second, "explanatory stage", they return to the laws and the empirical results in **4-D** in order to explain them as consequences of the **5-D** world. The *explanans* is the **5-D** structure and the *explananda* are laws and facts in **4-D**. First I relate Kaluza's and Klein's theories to the literature on explanation *qua* unification described in the first chapter and argue for the correlation between unification and explanation, *pace* Morrison. Second, I show that Klein improved significantly upon Kaluza's theory especially in respect of their (relative) explanatory stores. I investigate in what sense Klein's theory gets the right explanation from its unificatory power and in what sense other intended explanations cannot be inferred from its unification. The main questions about unification and explanation in this chapter are:

[55] *What is explained by Kaluza's and Klein's theories?*

[56] *How do we separate explanations from mere consequences of the unificatory assumptions?*

[57] *How are explanatory power and unification linked in Kaluza and respectively in Klein?*

[58] *What kind of brute facts do Kaluza and Klein rely upon?*

In a broader context I am interested in ranking Kaluza-Klein theory among gauge theories and in showing that Klein portended the theory of gauge invariance developed much later (with the notable exception of Weyl's 1918 paper). At the end of Chapter 16 I answer some philosophical questions related to the interpretation of Kaluza and Klein theories and I discuss more general philosophical or metaphysical issues such as realism, existence, etc. in direct relation to spacetime theories in which dimensionality plays a crucial role. I argue from a philosophical point of view against the conclusion frequently expressed in the physics literature that Kaluza's and Klein's theories are in any major sense trivial, spurious or ad-hoc.

Chapter 12. Unification of physical interactions

Why do we attempt to unify two physical interactions? Long before the heyday of **GR**, there have been philosophical and metaphysical reasons to believe that different interactions that looked or acted differently could be brought together under one and the same representation. Many scientists have sought a unity of science based on the metaphor of "nature as an organism" in which parts are related and connected. Two physicists with philosophical preoccupations are relevant here: E Mach and H. Weyl. Mach's idea of economy of thought and Weyl's program of removing dualities from physics were two popular doctrines of the unity of physical interactions among physicists and philosophers of the first decades of the 20th century.

12.1. Economy and duality of representations

The first aim of unification is economy. Ernst Mach, who influenced Einstein and the whole **GR** community, was one of the first to suggest that in order to achieve its goal, science has to meet a criteria of strength and *simplicity*: "Science […] may be regarded as a minimal problem, consisting of the [most complete][*] possible presentment of facts with the least possible expenditure of thought" (Mach 1893 490). He added that:

> when it is a question of bringing into connection two adjacent departments [disciplines], each of which has been developed in its special way, the connection cannot be effected by means of the limited conceptions of a narrow special department [discipline]. By means of more general considerations, conceptions have to be created which shall be adequate for the wider domain (Mach 1959 313).

Based on observations on humans and other living creatures, Mach believed in an economy of nature. Science should aim toward economy and simplicity, but also toward a "stable" representation (not disturbed by "new occurrences"), as complete as possible.[177] If the whole universe is similar to an organism, then its parts are interconnected: matter and field, charge and mass, energy and action. According to Mach's philosophical holism, the global presence of matter causally determines all local inertial forces and the inertial property of bodies. Mach tried to express this reduction to matter in the well-known "Mach's principle".[178] For Mach, mechanical interactions are coupled as parts of the same universe, despite some misleading appearances. The sources of mechanical forces are masses and other form of matter.

The Machian economy of thought was echoed if not in its letter, then in its spirit, by several preeminent figures of **GR**. Unity of physical interactions was an ideal shared by most mathematicians and by the avant-garde of theoretical physicists in the early 20th century. Some physicists philosophically prone went further and discussed the economy of thought as removing *dualities*. Hermann Weyl was deeply preoccupied with the duality of known fields. For him, unity had aesthetic appeal and an epistemic advantage: a more unified picture of the world warranted a deeper understanding of the laws of nature; any duality hindered the progress of knowledge.[179]

Back then there were several dualities present within the most advanced theories in physics:

- Duality between matter and fields,

---

[177] In his philosophical writings, Mach sought a unification of the physical with the psychical, but his conclusion applies to the pure physical sciences as well.

[178] The principle "Mach's principle", named and used by Einstein reads: "the *g* field is completely determined through the masses of the bodies. As mass and energy are the same according to special relativity and as energy is formally captured by the T_{mn} tensor, one can say that the g field is conditioned and determined through the stress-energy tensor of the matter" (Einstein 1916 241).

[179] Witness that Weyl is quoted at the beginning of Kaluza's paper.

- Duality between field equations and dynamics,
- Duality between electromagnetism and gravitation.

The first is indeed a huge topic in itself, still unsolved even today. In the first years of relativity, Einstein, Grommer, Weyl, Mie, Lanczos, *i.a.* attempted several times to derive matter from fields. In almost all respects, these attempts had failed and the question has been fundamentally reshaped after the advent of quantum mechanics. The second duality is also a complex problem that affects especially **GR** as a theory of the dynamics of fields in time. Bluntly put, the question of dynamics changes its meaning when "time" is not anymore a special parameter of the differential equations. We need to keep in mind that time is still special in **GR** because of the signature of the metric which privileges the direction of time. The whole discussion about the special regime of time is beyond the scope of this chapter.[180] Also there is no way to solve any of these dualities within a classical theory of field. The second duality may also involve a stable and mature theory of quantum gravity, which we do not have yet. I will focus especially on the latter duality, the one that is at the core of Kaluza's approach.

Weyl's idea of gauge invariance that foreshadowed the major developments in elementary particle physics after World War II is related to the idea of unification and symmetry.[181] The last sections of his *Space-Time-Matter* are dedicated to the idea of unity of forces and the way in which he identifies the "distance-curvature" with the electromagnetic field tensor F_{mn} as in the preface he decried some of the lingering dualities in Einstein's **GR** by noticing that in **GR** electricity and gravitation, "field" and "matter",

[180] See (Callender 2008) for an argument showing why time is special. Time in **GR** is still subject of controversies; see the discussion about time-translation invariance in **GR** in (Belot 2005, 255-291).
[181] See (Weyl 1970 sect. 35). Although the attempt from the 1918 book was a failure, Weyl aimed to take a further step in his 1928 book.

remain isolated one from the other. For Weyl, contrary to what other physicists suggested, the development of **GR** was not complete with these dualities at its core: "While the gravitation potential consists of an invariant quadratic differential form, electromagnetic phenomena are governed by a four-potential […] By so far the two classes of phenomena, gravitation and electricity, stand side by side, the one separated from the other." (Lorentz and others 1952 202; Weyl 1918 466). Weyl's proposal was to give a geometrical meaning of all physical quantities. His solution was not a reductionist one. He proposed that in order to unify **GR** and **EM** one need to create a different framework, in this case a new geometry, the "world geometry". Once this is created, we cannot in general make any arbitrary separation of electricity from gravitation. His theory can even help us to "comprehend why the world has four dimensions", which can be judged now as promissory at best, as we do not know precisely how many dimensions the world has (Weyl 1918 467).

Later on, D. Hilbert wrote a report for the Lobachevsky Prize in which he praised Weyl for "coalescing [*verschmeltzen*] in an organic unity electromagnetism and gravity"; quoted in (Scholz 2001 23). Weyl's method was based on the most general infinitesimal geometry known back then, the "projective geometry" developed by an abstraction from the affine geometry and on conformal geometry abstracted from Riemannian geometry. The metaphor in Weyl's approach is not elimination, but the organic (synthetic) unity. This "removal of dualities" did not work in the form envisaged by Weyl. We know now that he was wrong. The equation for the gravitational field was a differential equation of fourth order and not of second order, as expected. According to Einstein, both these geometries conflict with experience and seemingly the whole formalism does not describe the real world.[182] Albeit a false theory, it impacted tremendously the development of gauge theories in the following decades. Indeed, the infinitesimal “metric” geometry had a great theoretical appeal and foreshadowed the idea of gauge invariance of non-Abelian fields of Yang and Mills as well the theories of fundamental interactions.

Albeit strictly speaking false, I take Weyl’s program to be in the *spirit* of the philosophical ideal of unification of the known forces, not in its letter. His program was geometrical in its essence. He also thought of removing dualities without reducing a interaction to the other. He freed himself progressively from Mie’s program of *reducing* gravitation to an aspect of the electromagnetic field and moved toward a *synthetic* program that inferred both gravitation and electromagnetism from a more general geometrical approach, later baptized the “gauge theory”. By identifications, he managed to confer physical significance to several quantities of his formalism. He decided to identify the length curvature with the “electromagnetic” field tensor. He was able to read off the complete structure of Maxwell’s theory from this gauge invariance by identifications; see (56).

12.2. Unification and physical interactions

Both Weyl’s and Mach’s arguments can be taken as reasons to unify our representations of the world. My unit of analysis here is the scientific theory, more precisely the syntactic view of scientific theories. I talked about theories throughout this dissertation and here it is time to address the question: what is a theory? Bluntly speaking, I take theories as

[182] See Section 8.5 for Weyl’s theory. Weyl abandoned this idea in his 1929 book because of technical difficulties.

being set of sentences or hypotheses which are at least consistent.[183] As suggested before, the semantic view of theories would not drastically change the terms of my discussion.

Unification operates on two theories and outputs another theory. According to the definition **Def 1** (p. 12), unification of two theories T_1 and T_2, both describing two different classes of phenomena, is based on creating a new, composite representation of the phenomena within a new theory T. It is time to emphasize here the epistemic character of my approach; in discussing unification I start from representations of the world, not from the world "as it is". In a modest reading, that's where I start and that's where I end up. The success of unification is reached when the third representation is richer and more powerful than the previous ones—i.e. compared to their conjunction $T_1\&T_2$. I take explanation as one of the main improvements to be sought in this third, new theory. If other non-empirical virtues are going to be part of this package, so much the better for T.

Theories in physics at the beginning of the 20^th^ century. What is specific to the theories that attempt to unify two physical *interactions*? I need to particularize it to the specific context of physical interaction. What if the phenomena are physical interactions, for example, forces? Newton's and Maxwell's unifications both fall under this category. The electroweak unification and **GUT** are all unifications of physical interactions, each of them having certain problems and certain interesting features.

---

[183] There is a minimal condition in which we can talk about scientific knowledge. A scientific theory has to be internally consistent: we do not want to infer *p* and ~*p* from the same theory and the same interpretation of it. Of course, for different interpretations we may expect some severe inconsistencies, as illustrated by the interpretations of **QM**. Second, you want to have a consistency *with* other, well established theories. Non-locality of **QM** is a form of inconsistency *with* **SR**. In several cases, people are willing to give up or on the contrary to add something to the **QM** formalism in order to obtain Lorentz invariance. There are other cases even in the classical domain: **EM** theory in its Maxwell formulation was not consistent with classical mechanics. In cases like this we usually suspect one of the theories as being false. Indeed this is the intuitive approach to inconsistencies. In some cases, the formalism itself can give us logical inconsistencies. It is for example at the core of a recent debate whether **EM** itself is consistent (Frisch 2005, 212). For the present purposes, I take both **EM** and **GR** as internally consistent theories. Both theories are not consistent with **QM**.

I suppose that many attempts to unify several field theories have illustrated the economy of thought praised by Mach. Let us narrow the context of physical theories and see how physical theories have interacted in the first decade of the 20th century. One can see that right after the advent of **QM** (1926-1930), there were several theories in physics that were competing for a complete and economical representation of the physical world:[184]

Theories of physical interactions

- **Theories of bodies and their dynamics**:
    - Classical mechanics (**CM**)
    - Statistical mechanics (**SM**)
    - Thermodynamics (**TH**)
    - Special Relativity (**SR**)
    - Quantum Mechanics (**QM**)
    - Continuum (fluid) mechanics (**FM**)
- **Theories of fields and their dynamics:**
    - Electromagnetic Field Theory (**EM**)
    - General Relativity (**GR**)

[184] Some could say that out of my list some entries are "formalisms", not proper theories, because they do not have an interpretation in themselves. For example, **QM** has a different role as it can be considered a formalism in which other theories can be expressed. One can have a quantized theory T: **EM** was the first theory to be quantized with the Quantum Electrodynamics (**QED**), then the quantum chromo-dynamics etc. Maybe **SM** plays a similar role: it is apply to systems that are not related. In each case, an interpretation is necessary. Or maybe with each application **SM** is reinterpreted and restated. Then one should differentiate questions about formalisms and theories. I take both **EM** and **GR** as theories and not formalisms and I deal with the relation between **GR** and **EM**, in their classical, non-quantized formulation, specific for the first decades of the 20th century.

It is obvious that besides developing each of these theories we may be concerned with the relations among them. Philosophers and physicists have dealt with the possible relations among these theories. Setting aside the extreme positions according to which either there are no relations at all or, on the contrary, all these theories are appearances of a single Theory of Everything (**TOE**), I see several possible attitudes worth mentioning here: reduction, pluralism, emergence and finally, unification. In this chapter, I discuss in greater details the first two, before paying attention to unification.

There are some general features of scientific theories that describe physical interactions are easier and more specifically characterized. Here are some of the relevant features of theories of physical interactions:

- Physical interactions can be expressed as **PDE** or **ODE** having time as a parameter;
- They are local or on the contrary global in spacetime;
- They need boundary conditions
- They use potentials
- They have a invariance group to coordinate transformation

For classical fields, their representations are usually systems of equations in which fields are variables and the solutions provide their values at different space and time points. Sometimes these solutions are nothing more than *local* results. Sometimes they can be extended to any point of space and time and they become *global*. **EM** theory—as formulated in Minkowski spacetime—represents local values of fields at different moments of time. **GR** is manifestly a local theory and its results are in general not the same everywhere. On the contrary, **SR** is a global theory as it postulates the same form of the metric $\eta_{\mu\nu}$

everywhere. Similarly, the assumptions of Kaluza and Klein (**CYL** and **COMP**) impose global conditions on two "local" theories: **GR** and **EM**.

But there is more to be added to spacetime theories. Even for simple cases, the system of equations is not enough: boundary conditions as well as other specifications regarding global features of spacetime dictate what the solutions are. We also need to keep an eye on what symmetries can tell us about unification and explanation. The point is that in the case of physical interactions we need to specify more about this internal structure of theories—even for cases like **EM** and **GR**.

EM and **GR** are based on potentials. I already mentioned that the form of the potential plays a specific role (for the case of **EM** and **GR**, this is the Poisson equation). Roughly speaking, this gives the distribution of the potential in space. Besides the form of the potential, the two dynamics of the interactions matter.

In respect of unification, the new theory T has to provide the right dynamics and the right boundary conditions for the previous theories T_1 and T_2. If T_1 describes a field F_1 and its dynamics D_1 and boundary conditions B_1, and a theory T_2 describes field T_2 and its dynamics and boundary conditions B_2, we can envisage the unified theory T that describes an encompassing field F having components (or parts) F_1 and F_2. Theory T obeys a dynamics D which is at least *compatible* with D_1 and D_2 or, stronger, it is *derived* from D_1 and D_2. Then the boundary conditions B_1 and B_2 need to be integrated in the boundary condition B of T. These conditions of compatibility and integration can be done only in some specific conditions of specific dynamics; it is clear that when gravitation is involved we have to read cautiously into the dynamics of fields. The spacetime as a field is involved here, but in a very deep and fundamental way. Ditto for boundary conditions—because we deal here with partial differential equations instead of ordinary differential equations. Boundary conditions and sets of equations, together with other specifications that need to be explicitly stated form altogether a theory. If the unified theory *T* does not provide the right boundary conditions, then unification is in jeopardy. Last but not least, the way various mathematical structures (mainly tensors, but vectors and scalars, too) are invariant (or on the contrary are not invariant) to coordinate changes are also essential features of theories that describes interactions.

12.3. Inter-theoretical relations

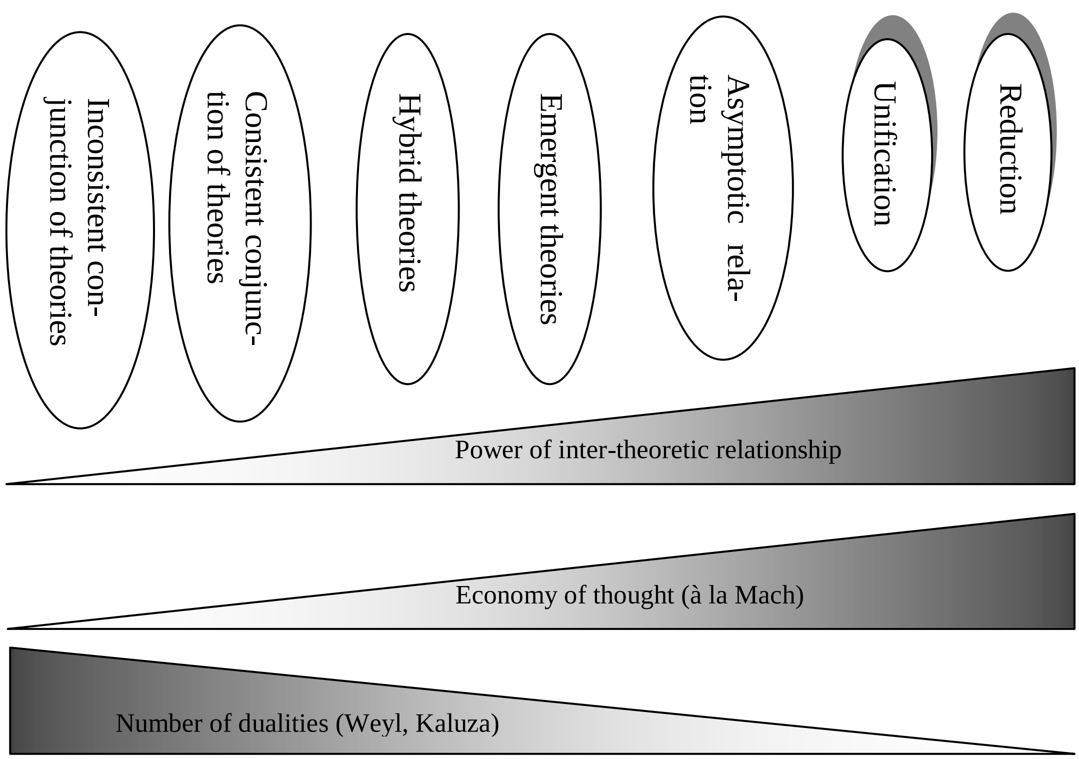

Figure 2 Types of interactions among theories

What are the alternatives to unification of two theories describing physical interactions? I see here two extremes: ignore completely the inter-theoretical relations and do not care even about inconsistencies. The other extreme is to reduce one theory to the other.

If one still wants dis-unity one can try consistency and adopt pluralism as a solution of a peaceful coexistence of the two theories: keep them dis-unified, i.e. take them as separate representations of the world, but be sure they are at least consistent.

I discuss here pluralism and reductionism. The latter is the "contrast class" to unification. I argue that in the case of **EM** and **GR** unification is an alternative to reduction because reduction cannot occur in this case.

Inconsistencies and unification. Before discussing reduction I take a look at the other extreme of the spectrum depicted above. There is a connection between the ideal of unification and consistency. Inconsistency is worse than an ugly theory or ever worse than a false theory. I associate the fear of internal inconsistency to a specific way of criticizing any unificatory program. Once we have unified two theories we may stumble upon unexpected inconsistencies within the new theory T, even if it was not originally present in the conjunction of T_1 and T_2. This is a very unfortunate case. In this case the culprit is the formalism of T which introduced surplus structure or has elements which are not desirable. The dis-unifiers will always seek inconsistencies within the formalism of the unificatory theory T. According to this dis-unificatory strategy, when inconsistencies occur, we need to go back and dismiss unification altogether. Such a radical strategy, dimly suggested by Morrison or by other dis-unifiers is not a clever strategy. Unification can reveal aspects which were hidden in the previous non-unified theories.

In the case of the Kaluza-Klein theory the problem of inconsistencies arises in the sense that Kaluza's theory does not apply to microparticles the same way it applies to the macroscopic objects. For Kaluza and Klein, we cannot have two physics: one for the micro world and one for the macro world. This is more than an incompatibility between two theories, but reveals an internal inconsistency: the theory cannot represent the object in spacetime using the same formalism. A different interpretation is necessary in order to apply the theory to macroscopic particles or micro-particles. This is a form of content-sensitivity suggested by Frisch according to which we need to add "rules guiding the selective application of the theory's basic equations" (Frisch 2005 193). When the rules are not present, the whole interpretation of the theory is open to arbitrariness.

As we will see shortly, T. Maudlin takes consistencies as a sufficient condition for unification (See section 12.6). But it is not clear whether we really achieved it even in the case of theories which are in themselves successful and explanatory.

I do not want to emphasize the troubles of the consistency condition here because even if we achieve internal consistency of the unificatory theory we want to look back at the main aims of unifying two theories. I suggest that even when we do not achieve perfect consistency, we may have reasons to move forward instead of dismissing completely unification.

Pluralism. Let us focus now on the peaceful coexistence once consistency is achieved. Why should one bother about relations between fields when one can be simply a pluralist and let all forces be described by different, proprietary theories? Such arguments of "let all the flowers bloom" type can be easily generated for almost any pair of theories. The folklore has it that Pauli once said that "What God hath put asunder no man shall even join".[185] How do we know what is set asunder and what not? The obvious problem with a deflated pluralism is that inconsistencies may arise among explanations and predictions. For example, the trajectory of charged particles in an **EM** field is not a trivial problem for **GR**. In **GR** one need to be able to account for such data.

The pluralist does not deny all inter-theoretical relations, but adopts a minimalist attitude. Inter-theoretical relations do not explain, do not predict and matter less for the current practice of science. Methodologically, one should proceed in a Cartesian matter and divide big problems in small problems and try to solve them piece by piece. In fact **QM** is the theory of small scale object, while **GR** deals with stars and galaxies. **CM** deals with low speed motion, while special relativity effects are relevant only to high energy levels. **TH** does not deal with individual particles, while **CM** is a theory of individual particles. **FM** is a theory with no particles at all in which continuum (mainly fluids or gases) replaces the particles. But the pluralist will admit that **FM** is closer to the theories of fields because both **GR** and **EM** postulate continuous fields (or weaker, they avoided dealing with and attributing explanatory powers to discontinuous fields). Electric and magnetic fields seem to exist independently of gravitation because they relate to charged particles and not to massive bodies.

For all practical matters, the pluralist strategy may work seamlessly. Situations when we need to deal with large scale phenomena and small object at the same time are rare. Cosmological models of the Universe and black holes may be some of the few cases. But in other cases it can in fact harm more than helps. Think of extreme situations in which

[185] Quoted in (Moriyasu 1983 102). It is not clear whether this was directed against any attempt at unification or a specific model.

one wants to know whether an intervention on E-B fields deforms *enough* the spacetime manifold. If **GR** is a theory about spacetime, then one wants to know whether high energy EM fields create a wrap in the fabric of spacetime. Who would like to know this? For example, the interventionist who wants to see whether there is a possible causal connection between **EM** and **GR**. Or let us say we are able to smash particles in a collider and create colossal fields **E-B** or whatever for short periods of time. Is this going to wrinkle significantly the texture of spacetime to create a black hole for example? Or maybe even a more mundane case would be the Earth's imminent collision with an asteroid. Somebody proposes to send an **EM** bomb in the proximity of the asteroid and detonate it. How strong should it be in order to fend the asteroid off? How close should we detonate it?[186] etc. If one thinks that **EM** pulses have nothing to do with gravitational fields then there is no answer to such questions. Too permissive an attitude would not provide answers to such questions. Inter-theoretical relations such as the one discussed in Kaluza and Klein can have immediate impact on our everyday life even if given the huge scale difference between the two theories, these are not accessible to our everyday instruments and may be well beyond the scope of actual technology.

Emergentism. In this dissertation I do not touch emergentism or the case of hybrid theories because I do not find it relevant to my case study. Here, the creation of the new, unificatory theory *T* is not a case of emergence, at least not in the standard meaning of emergence. In my view there is also an intermediate case, the so-called "hybrid theories" which are less than unificatory and closer to emergence: they are more or less conjunctions

[186] I do not claim that such a possibility is in fact relevant, but it is good to know that it cannot be excluded on *a priori* grounds.

of theories. It is relevant to ask whether the new theory *T* is really new or can be reduced to T_1 or T_2, to their conjunction, or some suitably modified forms of them.

12.4. Reductionism: electromagnetism and gravitation

At a first take, the answer to [54] (p. 262) is that the aim of both Kaluza and Klein was to unify the theories representing two physical interactions: the gravitational and the electromagnetic interactions.

We already face here a possible trouble: gravitation can be interpreted as an interaction or on the contrary it is not after all an interaction—or not in a straightforward way. Physicists and philosophers toiled to show that gravitation is more than an interaction or stronger, that it is not an interaction at all; it is the theory of space-time itself, not the theory of an interaction mediated by a force.

If gravitation is more than a force, or weaker, it is not an interaction in a straightforward way, then why should we unify it with other interactions or forces? If gravitation is more fundamental than the **EM** interaction or other interactions for that matter, then we should try to reduce everything to gravitation? Indeed, this is the substance of the strong geometrization program.

Inter-theoretical reduction is popular in areas such as philosophy of mind, philosophy of biology, even in philosophy of physics. How do we apply reduction when it comes to physical interactions? Successful reductions frequently cited are optics being reduced to electromagnetism and friction being reduced to electromagnetism. Sometimes interactions are too different to be reduced and unification can be the alternative to reduction. How do we unify two things which are different? This constitutes one of the difficulties that linger at the core of any attempt of unifying gravitation with other forces and I show in what sense Kaluza-Klein theory addresses this foundational issue. As discussed before, we unify representations of these two different "things", i.e. two interactions. To anticipate: in analyzing failed cases of reduction among physical interactions, one should look at the group of invariants associated to theories. As I indicated before, general answers to these questions are not useful here and we need to enter into the details for both **EM** and **GR**.

A "theory of everything" **TOE**, whatever it means, is in principle an attempt to reduce all physical theories to one, all-encompassing theory although it embraces explanatory pluralism.

CM can be reduced to **SR**, and that in some conditions **TH** is reducible to **SM**. When we look at physical interactions, reduction provides good explanations. Chemical bounds are electromagnetic forces, although in the majority of cases we need a quantum theory of electromagnetism in order to reduce them (many argue that in fact for atoms more complex than the hydrogen even this explanatory reduction fails for computational reasons). Collisions are reducible to electromagnetic interactions at the atomic level. Friction, despite all appearances is nothing more than an electromagnetic interaction. It seemed that for the vast majority of the macroscopic physics, reduction to electromagnetic interactions was germane in understanding a large pool of interactions—with one notable exception: gravitation. In other words, the **EM** theory has a special place in explaining the microscopic world.

The reductive attempts were preeminent in the relativity years: theoretical reduction was the sought relation among theories. There is an obvious way to relate **CM** to **SR** and further **SR** to **GR** and without further ado I can claim that **SR** is a special case of **GR** theory. At the beginning, **SR** guided the development of **GR**. Later on, **GR** had its own development and facilitated the discovery of phenomena well beyond what **SR** could predict: gravitational waves, black holes *i.a.* which were antithetical to **SR**.

There is a sense in which scale and size matters even in the case of **GR** and **EM**, both being long-range forces that act on bodies no matter what their size is. Gravitational effects are minor at the microscopic level, but relatively powerful in large-scale phenomena.[187] The macro-level to micro-level reduction was not promising at all: thinking of gravitation as a macro-level aspect of some electro-magnetic micro-level interactions turned out to be unscientific and very problematic. And vice-versa. Remember that both theories have non-trivial vacuum solutions: even in the absence of sources, the field can exist in the form of **EM** waves for example.

The new wave of reductionism. Although the "level" reductionism is not promising at all in our case, reduction is not completely dismissed. Reduction is not always couched in terms of micro- and macro-levels. Some reductions are between theories, like the reduction of **CM** to **SR**. Other reductions, although can be formulated at two levels, are still enshrined in mystery, for example the relation between **QM** and **GR**. In this case, which one is reduced, which one is the "reducing" theory? Taking one more fundamental than the other is a hopeless endeavor. Even in exemplar cases of reduction such as **TH** to **SM**, we need to revise heavily the model of reduction proposed by Nagel, Hempel, etc. Newer accounts of reduction, the so-called "new wave" (associated with Paul Churchland, Patricia Churchland and Cliff Hooker by (Endicott 1998; Endicott 2001)) make room for de-

[187] Large-scale electromagnetic phenomena were less frequent. The *Aurora Borealis* was one of the few large-scale electric and magnetic phenomena known in the 1900s. Sunlight is not charged so it is not influenced by Earth's magnetic field.

grees of reduction: “reduction may be smooth or bumpy, or anywhere in between” (Churchland 1979 84). There are also degrees of replacement, of old terms with new terms. Inspired by the reduction of folk psychology to neuroscience, the Churchlands think of reduction as an displacement of old theories, rather than a translation of old language into new languages. In some cases the reduction is elimination, in other cases the reduction is a major transformation of the old theory.

The old theory is corrected and sometimes the correction creates a different theory. In other words, if T_B is a new reducing theory and T_R is the old reduced theory, we create a new theory T_R^* which is deduced from T_B. Simply put, we create a new theory **CM*** which is analogue to **CM** but it is deduced (logically) from **SR** by limiting assumptions (usually called “conditions of reduction”, C_R (Churchland 1989, 321; Churchland 1986 288-290; Hooker 1981 49). One can tell that **CM** is displaced by **SR**. More formally (Hooker 1981 49):

$$\left(\left(T_B \,\&\, C_R \supset T_R^*\right) \& \left(T_R^* \mathbf{A_R} T_R\right)\right) \xrightarrow{\textit{warrants}} \left(T_B \mathbf{R} T_R\right) \tag{132}$$

where the $\mathbf{A_R}$ is the analogy relation, **R** is the reduction relation and $\supset$ is the deductive implication. The “bumpy” reductions are where the analogy $\mathbf{A_R}$ is a very weak one. The strong reductions are such that T_R^* and T_R are virtually identical (T_R^* is the exact equipotent isomorphic image of T_R) (Bickle 1992 417).

The new-wave reductionists warn us that not all derivations are reductions. The very idea of successor of a theory is sometimes misleading. Hooker even mentioned Kaluza: “Would Kaluza’s 5-dimensional unified electro-magneto-gravitational theory be a candidate successor to classical mechanics?” without providing an answer to this question

(Hooker 1981 44). In fact, his hand-waving remark of Kaluza leaves room for a non-reductive interpretation.

Asymptotics. We can show in some cases that a theory is the limit of a different theory. Typically we take a parameter of one theory and hypothesize what would happen when it has a critical value: zero, infinity or other singular value. Some popular, albeit contested, ways to perform such operations were a specific part of **CM** as a limit of **SR** and **CM** as a limit of **QM**:

$$\lim_{c \to \infty}\left(\mathbf{SR}\right) = \mathbf{CM}_1 \tag{133}$$

Before **SR** was discovered, the speed of light was literally only the speed of electromagnetic waves and **EM** was not related to mechanics at all. In **CM** or **SM** it made sense to ask questions such as: "how much energy do we need to accelerate this baseball to 300,001 km/s?" or "what is the probability that exactly two molecules of this gas could have speed between 300,001 km/s and 300,002 km/s in a given interval of time and given certain conditions?"[188] One can see that electromagnetism and classical mechanics acted somehow independently: **EM** waves could not propagate with a speed above the speed of light, although any other massive particle could. The other cases of "take to the limit", the correspondence principle, is less popular nowadays:

$$\lim_{\hbar \to 0}\left(\mathbf{QM}\right) = \mathbf{CM}_2 \tag{134}$$

where $\mathbf{CM_2}$ is a specific type of classical theory of mechanics. Another case discussed in detail by Batterman is:

$$\lim_{\lambda \to 0}\left(\text{Wave Optics}\right) = \text{Ray Optics} \tag{135}$$

[188] The speed of light is less than 300,000 km/s.

Let us try to apply asymptotics to our case study. A Unified Field Theory (**UFT**) such that:

$$\begin{cases} \lim\limits_{\substack{\text{some}\\\text{situations}}} (\mathbf{UFT}_{\mathit{fundamental}}) = \mathbf{GR}_{\text{coarse grained}} \\ \lim\limits_{\substack{\text{some}\\\text{other}\\\text{situations}}} (\mathbf{UFT}_{\mathit{fundamental}}) = \mathbf{EM}_{\text{coarse grained}} \end{cases} \tag{136}$$

gives us an intuitive sense of unification, at least. We create a more fundamental theory which, when it is taken to the limit, gives us two course grain theories, **EM** and **GR**. Some physicists shared the feeling that both **GR** and **EM** were coarse grained in the sense that they were in need of a better description. Both were mature, successful theories but both were stages of a X theory, difficult to foresee. One step taken in the 1940s was to quantize **EM** theory and this definitively improved its power of prediction. In respect of **GR**, the revolutionary changes were less impressive. The theory resisted several attempts to be quantized or changed in any significant way.

If we take Batterman's suggestion seriously, one can see **GR** and **EM** as idealizations or simplifications of still deeper theories. Removing the dualities mentioned and explaining better the interaction with quantum fields is part of this refinement that both **EM** and **GR** needed. The other two reductive options speculated in the 1920s:

$$\begin{cases} \lim\limits_{\substack{\text{some}\\\text{situations}}} (\mathbf{GR}) = \mathbf{EM} \\ \lim\limits_{\substack{\text{some}\\\text{other}\\\text{situations}}} (\mathbf{EM}) = \mathbf{GR} \end{cases} \tag{137}$$

were doomed to fail. The usual problem with such limiting cases is that the match is not perfect, especially in the case of (140). There are situations in which the very operation of taken to the limit one theory creates a third theory.[189] One needs to acknowledge the existence of theories that appear at the limit between two theories T_1 and T_2 as a result of idealizations, in many cases non-Galilean idealizations.[190] But the existence of an intermediate theory hinges upon some similarities between T_1 and T_2. We know that despite some similarities, the two interactions are essentially different. Mie and Nordström attempted to offer an electromagnetic theory of gravitation like a limit theory, but both attempts failed.

I conclude here that one should not be happy with the extremes of the possibilities depicted in Figure 2 (p. **Error! Bookmark not defined.**).

Are there chances to ascertain a reduction between **EM** and **GR**? How "bumpy" is the reduction of **EM** to **GR** or the other one, of **GR** to **EM**? I show here that both reductions are conceptually and practically impossible—or at least very inconvenient. Let us go into the details of the reductive relation between **EM** and **GR**.

There is no way to imagine that **GR** is the micro-level and the **EM** is the macro-level. In fact, gravitation is so weak a force that it does not matter that much at the micro-level—except the perturbation of the atomic model by the gravitational force which is absolutely negligible. At the scale accessible to us electromagnetic force is dominant and it is dominant to atomic scale, too. For normal energies and scales larger than the Planck scale, gravitation does not exist for the Standard Model of Elementary particles. Yes, there is a nice peaceful coexistence between the theory of matter and the theory of gravitation—for many practical purposes. Gravitation does not reign over the world of low scale

[189] Some cases discussed in the literature are phase transitions and catastrophe optics. See Batterman 2002 for details.
[190] R. Batterman and A. Wayne (private communication) discuss cases of such theories: the phase transition or the catastrophe optics. See (Batterman 2005, 225-244).

phenomena. At very small scales, at the Planck scale the situation may reverse radically: gravitation in fact becomes stronger than other forces and is an important factor in describing the world. But for the time being there is no such theory of gravitation at Planck scale.

So there is no way to continue to talk about a micro-macro reductionism, but a form of reduction in which the relation to a fundamental entity is preeminent. The theory closer to the foundation is the theory with higher chances to reduce the other one. What is then fundamental in the inter-theoretical relation between **GR** and **EM** and which theory come out as more fundamental? Not the levels or scale or forces and not levels of energy, but the spacetime structure is the fundamental level and **GR** is the theory that deals with it, unlike **EM** theory.

The electromagnetic program as a failed reduction. The "electromagnetic program", dating back to the end of the 19th century and early 20th century, was based on the assumption that the ether is described by the electromagnetic theory and that all laws of nature could be deduced from equations of the **EM** field; (see (Vizgin 1994 ch. 1; McCormmach 1970; Renn and Schemmel 2007 4:623-759) for a comprehensive historical description). In essence, this was a unificatory program, but it ran in the opposite direction than Einstein's. At different times and in different ways, J. J. Thompson's, G. Fitzgerald, O. Heaviside, M. Abraham, W. Kaufmann, G. Mie, etc. tried to infer mechanical properties of matter from the properties of electromagnetic fields. The scope of this program was to describe a variety of interactions within one framework. Other task was to express the mass of the electron as an electromagnetic quantity and to show that matter is composed solely of electrons. Other properties including length and density were thought as being deducible from the electromagnetic field inside matter. For example, Wien tried in 1900 to infer gravitation from the electromagnetic field, and to conclude that gravity depends on the electromagnetic field. Lorentz had adopted a more careful position. Instead of pretending that his transformation are real and they affect space and time themselves, he took them as sheer aids in calculation. He disliked the disunity which electromagnetic program had reintroduced in the core of electromagnetism by treating light and matter composed of electrons on different footage (Morrison 2000 156). But he still relied on the assumption that contraction of length depended on the electronic and atomistic constitution of matter. In his 1904 paper, Lorentz admitted that "the proper relation between the forces and the accelerations will exist in the two cases, [a system with translation and another without translation], if we suppose that the masses of all particles are influenced by a translation to the same degree as the electromagnetic masses of the electrons" (Lorentz and others 1952 30).

G. Mie developed the electromagnetic program, starting from Wien's and Abraham's worldviews, well over the borders of electromagnetism and gravitation. He used a variational principle and a world function in order to infer electromagnetism *and* gravitation.[191] Mie's approach was in sharp contradiction with Einstein and the period 1912-1916 is rich in exchange of papers full of reciprocal accusations of misunderstanding. Mie thought of **EM** as a theory about ether, more fundamental than gravitation. He hoped that all the properties of matter could be inferred from **EM**: spectra of atoms, mass and charge of electron etc. Electrons were the only elementary particles known in 1912, but Mie was

[191] Mie's main writings are: "Grundlagen einer Theorie der Materie" ("Foundations of a Theory of Matter" 1912), and "Bemerkungen zu der Einsteinschen Gravitationstheorie" (Remarks concerning Einstein's theory of Gravitation", 1914). All are translated in (Renn and Schemmel 2007).

already convinced that they are simply “knots” in the ether: a material particle is “a small region in the ether where the state variables take on enormously large values” (Renn and Schemmel 2007 655). The main desideratum was to have *one* system of equations whose solutions represent elementary particles where they are located and the Maxwell’s equations far away from the particles. The same system of equation could represent the gravitational field far away from its sources. The field equation of gravitation altogether could be derived from his “ether physics”. Needless to say, Mie’s other aim was to infer classical mechanics from the **EM** equations, too.

In pursuing this program, Mie went well beyond what Lorentz had said. Lorentz took electrons and non-electromagnetic forces as basic elements of the theory. Mie’s ambitions were to describe electrons and gravity as stable solutions of the **EM** field equation without introducing particles.

There are several novel approaches in Mie’s theory: he used a variational principle, he was maybe the first to suppose that a theory of gravitation has to be non-linear, and he correctly described the principle of a “theory of solitons”. But it was a fundamentally flawed theory in its reductionist ambitions. It seems that Mie’s unification was unphysical for several reasons. First, **EM** is not the fundamental theory to which gravitation can be reduced. Secondly, even in the first decade of the 20th century the quantum aspect of **EM** was experimentally proven. Mie did not incorporate this aspect in his theory. Second, he postulated a scalar potential such that the equations depend on its absolute value and not on the differences of its value as expected. In today’s parlance, his theory was not a gauge invariant theory. Third, Mie failed to find the “world function” *depending only on the field variable* that had to be added to the Lagrangian of **EM.**[192]

In the papers written soon afterwards, Einstein rejected almost all what Mie had supposed and reversed Lorentz’ procedure: instead of deriving mechanics from electrodynamics, he looked for those dynamical transformations that remove the internal inconsistencies of electrodynamics and inferred those mechanical transformations which removed the asymmetry between **E** and **B** and the subsequent disunities with dynamics (Morrison 2000 165)**.** In this sense, although **SR** heavily relies on **EM**, the dynamics of matter is not reduced to electromagnetic forces. Electromagnetism and any theory of matter do not explain length contraction or time dilation. Aside from the duality of the electric and magnetic fields, **SR** is taken to enact another more general unification: mechanics (including dynamics and kinematics) has been unified with the theory of the electromagnetic field—although not in the sense of **Def 8**.[193] Indeed, electromagnetism was what mainly gave the Lorentz transformations a full physical meaning.

We have now some examples of unificatory attempts prior to the dawn of the **GR**. The moral to be drawn is that “unification was in the air” and that several reductive attempts failed, either because they took the wrong direction of reduction or because reduction is simply not the name of the game. Last but not least, we saw a paradigmatic case of

---

[192] See Section 8.4sqq. for a description of Hilbert’s world function for **GR**.

[193] There is a very recent attempt to reverse Einstein’s procedure, but not in the sense of an electromagnetic program. In (Brown 2005), Brown defends an interpretation of SR in which geometry of 4-D spacetime cannot explain anything and it needs to be augmented by a dynamic understanding. Time dilation and length contraction are for Brown consequences of the microstructure of rods and clocks and not the result of the properties of the space-time structure. It is not the geometry that shrinks rods and dilates clocks. Their internal atomic structure, based mainly on electromagnetic forces, is the cause of such phenomena.

unification in **Def 8** which is too strong. I draw a map useful to situating my case study historically and conceptually. The next logical step is to look at **GR**.

12.5. The reduction of EM to GR

Some possible relations, from the weak coexistence to the strong reductionism, were already suggested by Einstein and Hilbert, among others. But there were very few hopes that **EM** and **GR** are related one to the other by idealizations or perturbations. Recall that the two forces were scales of magnitudes apart; gravitation is only attractive, electromagnetism is repulsive (for other differences, see Section 8.1). Nevertheless, for Einstein and the enthusiasts, reduction was supposed to do the job by expressing the electromagnetic quantities in terms of stress-energy tensor because *the* theory of the physical universe was **GR** and supposedly all other theories were to be reduced to it. Both theories claimed generality and universality. But there is something special about the claim of **GR** that it represents the world in the most general way. Indeed, according to Einstein, **GR** was *the* (only) theory of space-time and its interaction with matter and energy. **EM** was a theory about how a specific form of energy (that of charged particles and **EM** waves) is carried in spacetime. **EM** presupposed a background theory of spacetime; in **EM** the field does not affect spacetime, but lives in it. If **GR** is *the* theory of the dynamics of spacetime given the presence of energy, then logically **EM** is degraded to a theory of how charged particles and fields carry energy and nothing more. In this case, **EM** is part of the stress energy tensor $T_{\mu\nu}$ in the Einstein's field equation in the same way as any presence of matter (dust, stars) will affect the $g_{\mu\nu}$. According to such a reductive argument, **EM** does not play a special role and can be easily incorporated into **GR**. Electromagnetic sources are some of the possible sources of "warpage" of spacetime (Thorne 1994 117). It is also true that Hilbert proceeded in his derivation from Mie more than from Einstein so his attempt is more specific. In the revised version of the 1916 paper, he indeed averted from any similarity with Mie's theory. It is obvious that there are other sources for the curvature of the spacetime than **EM** fields.

What about the other relation? Are **EM** fields determined by gravity? Whether **EM** is totally determined by the curvature tensor has been a question under scrutiny since Hilbert. In fact, Hilbert and Rainich, *i.a.* adopted this reductive stance. Hilbert wrote to Einstein on 11/13/1915: "According to a general mathematical theorem, the electromagnetic equations (generalized Maxwell equations) appear as [a] consequence[s] of the gravitational equations, such that gravitation and electrodynamics are not really different" (Kox 1987vol8A-doc 140).[194]

First, now it is clear that Hilbert was wrong and that his "theorem" does not apply. It suffices to mention that some authors R. Geroch *i.a.,* have proven that for several spacetime structures, electromagnetism is *not* the consequence of curvature of spacetime (Geroch 1966, 147-187). More precisely, it does not follow uniquely from the 4-D geometry. Second, this reductive attitude has to face even a stronger argument nowadays: if electromagnetism is reducible to gravitation, what about other forces? Are all forces reducible to gravitation *à la* Hilbert?[195] The suggestion of the more recent geometrody-

---

[194] This solution did not satisfy all physicists, Einstein included. See recent results in (Brading and Ryckman 2008 113).

[195] A side note is in order here. It would be unfair not mentioning here that another force, call it here the "nuclear force" was *logically* predicted in order to explain the existence of the nucleus (discovered in 1908-1909). A new potential or force was necessary to overcome the electric repulsion between protons within the nucleus which is about 10^{-15} the size of the whole atom. A model of the nuclear interaction was developed by H. Yukawa only by 1931 because this nuclear field was difficult to measure. The classical theory of this nuclear force did not provide any insights into the nature of the nucleon, so the theory cried for a quantization. In 1934 Yukawa already predicted that the nuclear interaction was carried by the "π mesons" which was a different way to postulate its quantization. Because unlike the electric interaction, nuclear force

namics (**GMD**) program advocated by John Wheeler is reductive in nature, too (Wheeler 1962, 334). **GMD** reduced matter to dynamical geometry of the spacetime structure. **GMD** program has scored some successes in reducing **EM** to the features of the $g_{\mu\nu}$ field but failed to reduce other types of fields to the metric. Philosophers questioned the **GMD** on several grounds: logical, physical and methodological (Gruenbaum 1973; Stein 1972; Earman 1972, 634-647).[196] Third, what about the quantization of all physical fields but gravity? It seems that the strong reductionism faces major hindrances and has to be dismissed. As I already discussed, the other approach, the electromagnetic program, promoted by G. Mie and M. Abraham, which attempted to reduce gravity to electromagnetism, had faced insurmountable difficulties, too.[197]

There is also a serious problem relating the initial conditions. Imagine a "**GR** fundamentalist" who thinks that she can put any kind of boundary conditions on any kind of Cauchy surface in order to initiate any kind of **EM** system. Or could she? Again, we need to be able to reproduce the class of initial conditions demanded by the **EM** theory in a **GR** way. For a class of problems in **EM**, this is in fact not possible. As a side note, for any reduction of an theory T_1 of a field F_1 to a theory T_2 of another field F_2 it is important to check that the initial conditions used in T_1 can be reproduced or generated by T_2. For **EM** fields this is very difficult for a simple reason: gravitation is attractive, **EM** force can be repulsive, too. Moreover, as we will see **EM** has a different kind of symmetry than **GR** has.

---

was a short-range potential and because it was quantized already, a possibility of unifying it with gravitation seemed even more far-flung than unifying **EM** with **GR**. The new Yukawa force, born less than a decade after the advent of **QM**, was in fact a major knock-out to unified field theories. In my dissertation I decided to focus on the classical development of the unification of interactions in which quantization of the field is not performed explicitly.

[196] An interesting question is whether the new quantum **GMD** advocated by Butterfield and Isham is unificatory or reductive in nature, but this would takes us too far from the present purposes.

[197] See Smeenk and Martin in (Renn and Schemmel 2007 623-631) for an introduction to Mie's program.

The argument from symmetry shows that no matter how we manipulate the mathematical formalism, we cannot create or destroy symmetries of theories at the level of boundary conditions.

12.6. T. Maudlin's ranking of unifications

I argue that when reduction is not possible, unification is the inter-theoretic relation to be sought among theories describing physical interactions. Unification is neither a mere reduction, nor a mere conjunction of two existing theories or the instance of an emergent theory. Neither of these alternatives is attractive. Conjunction is trivial; the reduction of macro-level to micro-level has notorious epistemic problems with explanations, simplicity and multiple instantiations.[198] Asymptotic analysis in the spirit of (142) in the case of Kaluza and Klein can shed light on some controversial aspects, but it is not a complete analysis from my point of view.[199]

I believe it is important to look at unification as being "beyond reductionism" or as an alternative to mere reductionism or mere emergentism. I do not want to provide here a general recipe, but I claim that some reductions are too bumpy and messy to be considered reductions anymore. A sketchy proposal is to replace the bumpiest reductions with unifications based on the fact that T_B is significantly modified by the reduction process. If such an alternative does the job one expects from reduction, then unification is worth analyzing.

Let us go now into the details of the unification of two theories of physical interactions.[200] I suggested that one reason why reduction cannot work is the symmetry group. Each theory comes with its own symmetry group and if they are different reduction is not

[198] Problems related to the epistemic reduction are discussed in (Putnam 1979, 457; Lange 2000, 348).
[199] The other alternatives looks only remotely related to what unification does.
[200] I follow here (Maudlin 1996), although Moriyasu has a more detailed approach in (Moriyasu 1983, 177).

an option. Once we know how to relate the symmetry group of one theory with the symmetry group of the other we achieve some unification. This is the idea behind Maudlin's ranking. Besides the mere reduction of some symmetries to other, more fundamental symmetry, there is a lot to say about relation between the symmetries of theories than reduction or identification. I mentioned above that two major goals of unification of classical fields are to unify different force fields and, respectively, to unify a force field with its source. In **SR**, the first goal can be achieved by identifying the electric and magnetic fields with components of the tensor field $F_{\mu\nu}$ such that a Lorentz transformation transforms the components of one into the other. The distinction between electric and magnetic fields disappears in relativistic electrodynamics: in the new **SR** ontology the frame-independent field tensor replaces electric and magnetic fields. We will see that such a mechanism of unification is only *partially* present in Kaluza-Klein theory.

Maudlin's suggestion is that symmetry can describe in a very systematic way the process of unification as a creation of a new theory with a new symmetry group. According to Maudlin, if the symmetry group is a direct product, the unification is "so-and-so", or incomplete. In a stronger reading, it also can be judged as arbitrary or ad-hoc (although Maudlin does not use *ad-hocness* as a property of theories). Maudlin's account of unification relates it to a mathematical formalism and more precisely to the symmetry of theories. In other words, unification can be subjected to an "argument from symmetry", maybe the most popular in theoretical physics in the last century. For some philosophers, symmetry is "the primary clue to the theoretically constructed world" (Van Fraassen 1989 216). The well-known argument based on symmetry is M. Gell-Man's discovery in 1962 of the omega-minus (Ω^{-1}) elementary particle. Gell-Mann postulated that strong and electromagnetic interaction can change one particle into the other only if the hypercharge is conserved. Here symmetry was used to predict two new particles, that would have been discovered later (Hon and Goldstein 2006 436).

Many agree that the symmetries of two gauge theories can give us a clue whether their unification is a mere conjunction or can provide a real unification. The ranks of unification help in judging the quality of the unification from the strength of its symmetry. In his attempt to rank the varieties of unification in theoretical physics, Maudlin imposed three conditions on any non-trivial unification of two theories T_1 and T_2:

[59] T_1 and T_2 have to be consistent,

[60] the field force in T_1 has to obey the same dynamics as the field force in T_2 and

[61] there is a lawful (or nomic) correlation among the forces described by T_1 and T_2.

Both [60] and [61] are necessary. In the case of **EM** and **GR** a model with gravitational force but no electromagnetic force (or vice versa) is still possible so [59] is met. The dynamics of the two theories is not exactly the same, but they are similar. According to [61], the connection or correlation among the unified forces in the **EM** case is given by the fact that the variations of the electromotive force produce magnetic forces. It is not trivial to see whether this is true or not for the Kaluza or Klein theories: does a variation in the g_{mn} give rise to a variation in (Maudlin 1996, 129-144) F_{mn} and viceversa?

The necessary conditions [59]-[61] constitute the lower limit of unification. At the other end of the spectrum, Maudlin situated two cases of “perfect unification”: the electrodynamics unification, as well as the unification of inertial and gravitational masses in

GR. "Perfect" unifications provide novel explanations: for example, **GR** provided predictions and explanations that have been confirmed much later. From here, an enthusiast for unification such as Friedman can make a further step and commit to realism by believing in the entities postulated by the unifying theory premised on a simple group. Otherwise, theories premised on composed group are too trivial and easily obtained for virtually any interaction.

Maudlin noticed that many gauge theories, praised as embodying unification, do not qualify as 'perfect'. For example, a trivial case of gauge unification is when two gauge theories T_1 with the symmetry group G_1 and neutral particle X_1 and, respectively, T_2 with G_2 and neutral particle X_2 are "pasted" into a product group $G_1 \otimes G_2$ without any further ado. The standard model itself was build up as the product group: $SU(3) \otimes SU(2) \otimes U(1)$. SU(3) (the color group) is the representation of the local symmetry whose gauging gives rise to quantum chromodynamics (**QCD**). The *SU(2)* gauge theory is the group of the weak interaction. *U(1)* is the symmetry group of EM theory (see Section 14.3 for more details). One problem is that the same unification can be achieved by other groups. For example $SU(3) \otimes SU(2) \otimes U(1)$ is completely contained in the subgroup $SU(3) \otimes SU(3) \otimes SU(3)$ of E_6 (Georgi 1999 308). Which one is the suitable group? The argument from symmetry can sometimes be very ambiguous. Other considerations did eventually show that E_6 is not suitable.

Another problem is that these 'pasting' unifications could be nothing more than conjunctions of various dynamics. A next level of unification can be achieved when the product gives rise to observable forces and observable particles created from *mixing* the groups G_1 and G_2 by a “mixing angle” between X_1 and X_2. In the case of the electroweak unification, the group is $SU(2)\otimes U(1)$. Even at this level, some physicists (H. Georgi, K. Moriyasu) suspect “a partial unification, at best” (Moriyasu 1983 110). The upper level of gauge unification is premised on the simple gauge group (which is not decomposable in a product, as above). Grand unified theory (**GUT**), very popular about two decades ago was based on the idea that all interactions but gravity fit well into the simple group SU(5). But several other drawbacks of the theory made it less popular in the 1990s. There is a good sense that reduction among theories describing physical interactions is possible only when the group G_1 is a subgroup of G_2. Without entering here into details, it is easy to see that the group of **EM**, i.e. $U(1)$ is not a subgroup of the group of **GR-**arguably the diffeomorphism group $Diff(\mathcal{M})$ or vice-versa. Gravitation is not a standard gauge theory or taking it as a gauge theory is premised on several dubious assumptions.[201] Reduction again seems to be a hapless possibility. The peaceful coexistence of two theories premised by pluralism can be represented as a direct product, but such a possibility is not very attractive, either.[202]

In my case study, ranking Kaluza and Klein among gauge symmetries is a difficult task because gravity is not a gauge theory in a trivial sense as particles do not couple to the gravitational field; they *exist* in the spacetime. Even if primarily Kaluza-Klein theory is not a theory of interaction among particles and even if the gauge classification does not apply to this case, the Kaluza and Klein theories can be ranked accordingly. I see a major difference between Kaluza and Klein here, so a separate treatment is necessary.

---

[201] For a discussion see (Weinstein 1999, S146-S155; Belot 2003, 189-225; Redhead, Butterfield, and Pagonis 1999). It is true that this is the premise of the canonical quantization programme (thanks to C. Wüthrich for making this point clear to me).

[202] In the context of Klein’s theory there will be a more comprehensive discussion of this issue.

My main argument against putting Kaluza-Klein in Mauldin's schema is twofold. First, gravity is not obviously a gauge theory and more work is needed in this respect. Second, even if we take it as a gauge theory, with gravitation we need to add the boundary conditions to incorporate the boundary conditions needed by the **EM** theory. One boundary condition imposed upon Klein's theory refers to the fixed values that p^4 can take and the inherent interpretation of p^4 as related to the elementary electrical charge. From here one can see that a further condition needs to be added to the discussion of gauge theories in Mauldin: the boundary conditions of T_1 and T_2 has to be "the same" or at least consistent.

The moral I can draw from Mauldin's analysis is that to better understand unification, we need to look to actual scientific unified theories and to their details as no formal approach can separate trivial unification from real ones or exemplar unification. When we do analyze Kaluza and Klein, we'll see that both theories can be ranked between Feynman's totally trivial example and Mauldin's perfect unification. I argue that Kaluza and Klein are neither of them and that they both have some specific aspects not discussed in the literature, although in the development of Kaluza-Klein we will encounter echoes of these two poles. Mauldin's main result is to sketch an answer to the question:

[62] *Is unification between two physical interactions always possible?*

There are slim chances to answer this question in general. Maudlin suggests some necessary conditions of unification. Klein showed that the symmetry group of **EM** can be embedded in a specific 5-D type of gravitation. Later on, several authors showed that Klein result can be generalized to non-Abelian gauge theories, comprising a pretty general class of physical interactions. Partial and interesting answers to [62] in the context of spacetimes with extra spatial dimensions are also possible. The relevant theorem to be discussed is the

Campbell-Magaard theorem (see Section 13.1). When or why unification is not possible of course is again very context dependent. The only result I am concerned here with is the unification of **EM** and **GR** which is an interesting, albeit limited answer to [62].

The argument from symmetry in the case of unificatory theories is not sufficient. It acts as a sufficient condition only when we can specify the symmetry group of the representation of the two theories. Even so, the simplicity of the unificatory group of the theory T is in many cases at stake.

Chapter 13. Ingredients of Kaluza's unification

In Chapter 10 I showed that Kaluza managed to unify **EM** and **GR** in several steps. In order to achieve this goal, he needed to postulate a **5-D** manifold. Then he identified the suitable Christoffel symbols with the **EM** $F_{\mu\nu}$ tensor **(ID_1)**. Finally he modularized, i.e. divided in sectors, the **5-D** g_{mn} tensor in parts which each had a specific function. In this chapter I enter into the details of this procedure by keeping an eye on the details of the major elements of Kaluza's unification.

Kaluza started from the same desideratum to unification as Weyl did: give an expression of **EM** within the **GR** theory. There is some textual support to my claim. In the second paragraph of his paper, Kaluza hinted toward the economy and simplicity by citing and praising Weyl's surprisingly bold thrust toward the elimination of the dualism of gravity and electricity, "one of the great favorite ideas of the human spirit" (Appelquist, Chodos, and Freund 1987 61; Kaluza 1921 859). Similar to Weyl and to Hilbert, Kaluza would appeal to geometry, but in a radically different way. He endeavored to fulfill a "more perfect realization of unification" than Weyl's by having the source of gravity and electromagnetism fields stemming from a single "universal tensor": a tensor that packs in one mathematical form two types of interactions.. As we saw, some parts of this universal tensor were identified with gravity and other parts identified with electromagnetism. I offered an argument why this universal tensor has to be of a higher dimension.

Kaluza realized that is not easy to find the universal tensor. In general, "unification" should be used cautiously as we want to avoid an inflation of unification in science, including spurious unifications. Mere mathematical identities are not unification. I suggested already that tweaking components of tensors can integrate almost everything in tensors and that is not unification. Identifications by *fiat* can also have the same effect: they are not unifications either. Kaluza realized that contrary to what Hilbert thought, **EM** cannot be reduced to gravitation by simply writing the $F_{\mu\nu}$ within the $T_{\mu\nu}$ only. Equally, if one thinks that Einstein field equations unify, then all forms of energy that are expressible as a T tensor are "unified" with **GR**. A hodge-podge T tensor does not unify. Equally, force would be "unified" with acceleration in the second law of Newton $F=ma$, internal energy would be unified with heat and mechanical work in the first law of thermodynamics, etc. But it is not! Obviously, we do not want such a weak concept of unification and we do not want to get caught in Feynman's humorous trick. Unification is not a simple mathematical equality and any kind of mathematical operations like these should be taken with a grain of salt. Another unification can be achieved if one includes g and F in the same structure. The same can be said about the Lorentz group. A **5-D** Lorentz group can accommodate several 4-D Lorentz groups. Vectors in **5-D** can be regarded as **4-D** vectors of spin-1 to which a scalar of spin-0 is attached. A (0,2) symmetric tensor in **5-D** has 15 components that can accommodate a symmetric **4-D** (0,2) tensor with 10 components, a **4-D** spin-1 vector and a spinless scalar. If gravity is described by the **4-D** symmetric tensor g and **EM** by the 4-D vector A_{μ}, a theory in **5-D** has enough resources to represent **EM** and **GR** within the same **5-D** (0,2) symmetric tensor. Of course, other forces than **EM** would require extra dimensions to be represented.

The original idea of incorporating the electromagnetic field in the metric is astonishing and unique. However, at this stage at least, some features of the theory nourish skepticism. The operation of adding dimensions to tensors can be useful, but does it have any bearings on the world, i.e. is the extra dimension real? It's not easy to answer this question, even almost a century past Kaluza's approach. In subsequent chapters I focus on explanation, problem solving and *ad-hocness*. Here I am interested in discussing Kaluza's unification *per se*. Most preeminently, echoing the worries about unification in general (Morrison, Woodward, *i.a.*), one can see Kaluza's approach as a mere successful mathematical "notation" in which new components have been added to the metric g in order to geometrize **EM**. Here are the questions addressed in this chapter:

[63] Is Kaluza's unification by any means trivial, i.e. close to a conjunction of **EM** and **GR**?

[64] Is Kaluza's unification reductive or synthetic or does it constitute a third type of unification?

[65] What is the interpretation through symmetry of Kaluza's unification and what is Kaluza's place in Maudlin's ranking of unifications?

In respect of [63], I argue that there is unification in Kaluza theory and that it cannot be downgraded to a mere conjunction, although it is closer to a trivial unification when compared to Klein's theory. In answering [64], I show that Kaluza's unification is not reductive in the sense that **EM** is not reduced to a gravitational type of field in **5-D**, and that there is a synthetic element in Kaluza. As Kaluza's unification is not reductive, I show in what respect **5-D** gravity is different from the **4-D** gravity. I also show the limitations of Kaluza's theory in respect of [65]. In short, Kaluza's theory does not display the advantages of Klein's in respect of symmetries and being able to capture the group of **EM** as a part of the symmetry of the spacetime structure. In Maudlin's ranking scheme, I place

Kaluza's theory lower then Klein's. Moreover, in the following chapter I show the limitations of Maudlin's scheme when gravitation is involved.

13.1. The Campbell-Magaard theorem

There are two aspects of unification when we enrich the spatial structure with extra dimensions: one is the *embedding* on the physical 4-D structure into a **5-D** structure and it belongs to geometry. The second is the *interpretation* of physical quantities in the **5-D** structure and this is where physics becomes crucial.

Adding new dimensions to space or to spacetime is a much older endeavor. There are powerful results about n-dimensional spaces known to mathematicians as early as Newton. Very complicated curves and surfaces in lower-dimensional spaces can be unified through projections and reduced to simpler curves in higher dimensional spaces. Newton proved that all plane curves defined by polynomials of degree 3 with two unknown variables can be obtained as projective images of just five types of polynomials. The main results in n-dimensional geometry are discussed in Riemann's 1854 habilitation thesis. In the foreword of the thesis he added: "Abstract studies such as these allow one to observe relationships without being limited by narrow terms, and prevent traditional prejudices from inhibiting one's progress". Interestingly enough, "projective geometries", developed by A. Cayley and F. Klein, in which more than three numbers were associated to one point in a D=3 space was by far more successful than the interpretation of n-dimensional space as real directions.

Analytical mechanics, in its formulation of Euler, Lagrange and D'Alembert, had attempted to reduce dynamics of particles in 3-D + time to geometry of abstract spaces like configuration space or phase space. For a system of N particles, the phase space has $6N$ dimensions. Some of these dimensions can be associated to space dimensions, but not necessarily. Phase space is the space of representation: spacetime has a different meaning. Writing down the laws of physics in a geometrical way is one of physicists' oldest dreams (Hermann 1978 iv). Descartes was to first to suggest the geometrization of physics and in some respects analytical mechanics achieved his dream. For a mechanical system there are usually different equations describing its dynamics: a set of equations for positions, another set of equations for velocities, another for accelerations, etc. The reduction of number of equations describing the same system by incorporating different variables in one and the same vector is done in the configuration space which includes the spatial dimension q_i and the dynamical one as $\dot{q}_i$. The dynamics of real systems is a hyper-surface in an abstract space conveniently chosen trajectory. The topology of this abstract space was isomorphic to $\mathbb{R}^n$ because there were no reasons to do otherwise. Perhaps Riemann and Clifford were the first who had tried to use topology and higher dimensional space to simplify the laws of nature and to express everything in geometrical terms.

The major question not answered in the 19th century was to connect this multidimensional geometry to physics. Although the motion of a system of N particles with n degrees of freedom can be suitably described in analytical mechanics in phase space or configuration space, such spaces are purely representational and have no reality at all. The degrees of freedom a body has in a 3-D space and the multidimensional space of Riemann were not one and the same thing. The same system with a different constraint will follow a different trajectory in the phase space. The question to be asked is whether the *physical* space is embeddable in a n-dimensional manifold. It is also important to add some physical qualification to this question: giving the types of forces and objects we have in our physical world, is it embeddable in a higher dimensional space? The answer to this question did not come from Riemann's geometry or from the projective geometry: J. E. Campbell, a mathematician with interest in physics, proved in the 1920s that any *n*-space is surrounded by a vacuum *(n+1)* space (Campbell and Elliott 1926 212sqq.). In other words, a semi-Riemannian four-dimensional manifold is locally and isometrically embeddable into a five-dimensional Ricci-flat manifold. In its modified version by Magaard (1963), the Campbell–Magaard theorem states that it is always possible to embed—at least locally—solutions of the **4-D GR** in a **5-D** Ricci-flat manifold.[203] Another question asked by Campbell was this: how many extra dimensions are necessary to locally embed the *n*-dimensional Riemannian manifold in a higher dimensional space? According to Campbell's theorem, we need only one extra dimension to *n*. Campbell started from the Einstein field equations in *n+1* dimensions, split the metric and showed how to infer the metric of a n-dimensional universe with matter and fields from an empty universe of *n+1* dimensions. As a mathematical result, the Campbell-Magaard theorem has several assumptions which are difficult to reproduce here. But a succinct form of it is useful:[204]

> ***[66]*** *Any analytic Riemannian space* $\mathbb{V}^{n}$ *having a signature (s,t), i.e. with s space dimensions and t time dimensions (n=s+t) can be embedded* ***locally*** *in a Ricci-flat Riemannian space* $\mathbb{V}^{n+1}$ *with a signature (s+1,t)* ***OR*** *in a Ricci-flat Riemannian space* $\mathbb{V}^{n+1}$ *with a signature (s,t+1).*

---

[203] The rigorous proof came only in the 1960s in a PhD thesis by Maagard.
[204] P. Wesson, "The Meaning of Dimensions" in (Petkov 2007 9).

What is the relevance of this result? Campbell's result gives a mathematical justification of any attempt to embed gravitation in higher dimensional space. A mathematician can simply imagine our 4-D world as the boundary or the surface of a D>5, richer and more complex world. What this result says is in fact the opposite: the hyperspace is in fact simpler than the spacetime itself because it is not filled with matter—the wood in Einstein's metaphor. The *n+1* theory has R_{mn}=0.[205]

According to the Space-Time-Matter theory, promoted by P. Wesson *i.a.*, the Campbell-Maagard theorem is the mathematical basis of a program to infer matter in 4-D as described by Einstein's field equation (with $T_{\mu\nu}$) from the apparent vacuum in **5-D** described by an equation R_{mn}=0. The higher dimensional gravitation is flat, *i.e.* it does not contain matter and it is Riemannian, exactly what Kaluza-Klein would need.

Critics usually question the weakness of this piecemeal re-representation. It is not clear whether the embedding is a "*new* geometrical representation of field equations, which would be most likely what is required to have a useful geometrization of Unified Field Theories".[206] The same question can be asked about Kaluza-Klein theories: do they actually say something new about the **EM** equations? The argument here is more subtle: Klein theory indeed says something about the **EM** equation, while Kaluza says less about the **EM** field. In other words, we can ask whether a non-compactified theory can say something new about the **EM** field or any other 4-D field.[207] In my interpretation, Kaluza is a non-compactified theory, while Klein is a compactified theory and both are global

---

[205] This is emphasized by the Space-Time-Matter consortium (P. Wesson, P. DeLeon etc.).
[206] For a defense of the Campbell-Magaard theorem by Wesson and collaborators, see (Wesson 2007, 254; Wesson 2005; Seahra and Wesson 2003, 1321-1339)
[207] For example, Anderson argues against the so-called "non-compactified Kaluza-Klein theories" promoted by the Space-Time-Matter consortium (Anderson 2004 sect. 8.3).

theories. For the present purpose, I will argue in the next chapter that Klein says more about the **EM** field because he explicitly uses compactification. But Campbell-Magaard legitimates Kaluza's attempt, even if it historically came after Kaluza had published his paper.

Campbell's theorem is a local result and both Kaluza's and Klein's approaches are global (they postulated **CYL** and **COMP** everywhere). The two conditions in Kaluza and respectively in Klein can be deemed as "background dependent" so none of the theories qualifies as background independent theories. As claimed by many physicists (Smolin, Rovelli i.a), a background-independent theory leads to more elegant equations (Smolin 2005; Rovelli 2004).[208] It is difficult to say at this stage what role this condition plays in the Kaluza-Klein theory. In short, Kaluza and Klein theories are not background independent.

13.2. Step 1: The "make room" procedure

In order to see why Kaluza's result is not trivially a conjunction, I discuss his theory as a unification strategy. One can say that Kaluza "made room" for **EM** in the **5-D** manifold. It is literally a way of creating space in order to explain something that is not explainable in our 4-D spacetime (or alternatively in our 3-D space). It can work for different purposes, but in many cases it is a blatantly ad-hoc procedure because it can create ad-hoc explanations or predictions for almost anything (see Section 15.6). What is the schema of the "make room" procedure? We think or we suppose there is an X in the world (although

[208] This is a very controversial issue and I do not think Kaluza or Klein adds something to the whole discussion. A quick reply is that the very notion of signature is built into the theory, even into a background-independent theory. If signature is part of the background, one can think that the topology of the fifth dimension plays a similar role.

we may not observe or interact directly with X) and we hypothesize that X is only an aspect of a Y reality in extra dimension(s). On the contrary, Y can be as material as any ordinary object: our perception of Y is distorted. Or, if you prefer, the perspective we have on Y is wrong, incomplete or distorted.[209] If one wants to explain something else, let us say a scalar field such as the temperature, then one can add an extra dimension that makes room to temperature. In the vast majority of cases, we do not need such a new dimension either. Ditto for a temperature dimension, or any kind of dimension arbitrarily added. We have better explanations for ghosts, spiritualism, evil demons, temperature, etc. But we cannot rule out completely the procedure of "making room for X" for any X whatsoever. I am interested here especially in the case in which X is a collection of laws of physical interactions and the strategy is leveled to unify them.

In Kaluza's case, X is the unity of electromagnetism and gravitation and not a specific phenomenon. The *explananda*, as we will see later are laws, regularities or simply facts. But let us be liberal here: there are plenty of quantities X that can be part of this "make room for" strategy. The make room for can be interpreted literally: we create a space which has room for a specific quantity X.

The "make room" strategy for physical interactions was already suggested by Nordström: enlarge the structure by adding one extra dimension to it! Apparently, the new dimension is as spatial as the three other spatial dimensions. But there are several major differences that make this new dimension "special". Kaluza added it in order to create the conditions of unifying $F_{\mu\nu}$ and $g_{\mu\nu}$. This add-on is conservative in at least one respect: it is

---

[209] For example, for H. More and others X was the spiritual dimension so he added the "spissitude". One can now refute More's dimension on several grounds. See Section 15.6 for details.

equipped with a pseudo-Riemannian metric. Kaluza kept the gravitational field in **5-D** and mirrored *some* of the features of the 4-D spacetime: its metric, its signature, its zero Ricci-curvature: neither Kaluza nor Klein added matter fields in **5-D**. I want to investigate the "make room" strategy as used by Kaluza in order to achieve unification. The second step is crucial because it involves the identification of parts of the new **5-D** mathematical structure with structures in **4-D**, more precisely the laws and equations governing **4-D** electromagnetism.

13.3. Step 2: Identifications of Christoffel symbols

A further step after adding structure to the spacetime is to represent forces within this new structure by identifying its parts with physical quantities. I claim that this step was taken by Kaluza and Klein, although on a different pace, and perhaps in two different directions.

Kaluza's procedure is similar to other types of unification, especially to the **SR** unification. I claim that Kaluza illustrates well a type of unification based on *identifications*, used by Newton, Maxwell, Einstein, present also in electroweak unification or in String Theory.[210] In all these cases, the identifications brought about several philosophical problems.

Witness the identification of the fourth direction of spacetime with time in **SR**. It is interesting to discuss Kaluza's identifications in a different, but similar, context: the identifications used in the covariant formulation of electromagnetism. Minkowski showed how

[210] Newton used the second rule of philosophizing to identify multiple trajectories as caused by one and the same force: the same force pulls all the planets on their orbits around the Sun and the Moon around the Earth. Maxwell used identifications in order to unify **EM** and optics: "we can scarcely avoid the inference that light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena" (Maxwell (Clerk) 1861 21-22).

to transform a three dimensional space and a one-dimensional time, each with its own metric into a four-dimensional continuum with a metric that describes the behavior of photons and free material particles. In Einstein's 1905 paper **SR** has the features of a unificatory theory, although not the unification of physical interactions. In a nutshell, the general procedure to infer the 4-tensor of **EM** is by guessing its form from the equations written in three dimensions. For vectors, the element added on the fourth place is in many cases guessed and added by empirical considerations (the conservation of charge), or by theoretical, esthetical or computational, considerations. There was no rigorous standard procedure to infer the tensors of **EM**. Minkowski tried to equate time t with the x^0 coordinate of the 4-D manifold and his identification worked:

$$x^0 = ict \tag{138}$$

Here the = sign does not stand for a perfect identification. As it is, this equality is arbitrary from a physical point of view: other expressions could have been used as well. Although arbitrary the identification worked for all purposes in **SR**. Between 1913 and 1916, Einstein adopted the same methodology of "closing the circle" by guesswork: postulate X, write theory of X, and see how X kicks back. With "luck", X can fit the results one was looking for.[211] But we are still reluctant in identifying literally time with the complex number $\frac{x^0}{ic}$. But the form inferred by Minkowski worked in several contexts and the identification procedure solve problems. Many people still believe it is *ad-hoc*. It does not reveal something fundamental about the nature of space and time, or if it does, it is instrumental, at best: it reflects the way we measure space and time and the way we spa-

[211] See a recent talk by J. Norton about how Einstein inferred EFE by this method of "closing the circle" at PSA 2008.

tialize time. According to some interpretations of **SR**, time in the 4-vector **X** is not the same as the time measured by ordinary clocks, although they may be correlated.

In the context of **EM**, there is a deeper schism between what we observe and theoretical entities. In its covariant form, **EM** is unification by identifications, too. We identify **E** and **B** as parts of the tensor $F_{\mu\nu}$, which is covariant. If the ideal of modern theories is to find their covariant formulation, there we are: only the tensor formulation of EM is manifestly covariant. In modern physics covariant quantities are preferred to observable quantities. We write covariant equations in $F_{\mu\nu}$ and A_μ and we measure **E** and **B**. Is this a serious problem? The main questions a philosopher could ask is: how real are quantities like A_μ, $F_{\mu\nu}$ or even f_μ? (Healey 2004, 619-642; Leeds 1999, 606-627; Healey 2007, 297). G. Belot discussed three interpretations of the classical, i.e. non-quantized electromagnetism that can be succinctly put in this form: What are we supposed to take as physical, i.e real?: 1) A_μ, 2) ***E*** and ***B*** or 3) the holonomic interpretation in which *closed curves* in spacetime carry the electromagnetic properties, not points in space and time (Gordon Belot 1998 542-545).[212]

Identifications fit well the standard **D-N** model of reductive explanations where the theoretical identifications are the core of theoretical reductions. If light waves are identical to electromagnetic waves, there is a law-like correlation between the two. On one hand, the identification explains many optical phenomena as electromagnetic wave phenomena. In general, the reduced theory is often corrected and qualified after reduction. On the other

[212] The holonomic interpretation is related to the fiber bundle formalism and I do not discuss it in this thesis, although it is highly related to the Kaluza-Klein theory.

hand, the reducing theory does not survive intact the process of reduction.[213] Because I claim that Kaluza's identifications are not reductive, I can set this problem aside for now.

What about identification within unification? Inspired by the role identifications play in reduction, the role of mathematical identifications in unification has been documented in the literature on unification (Friedman, Morrison, Maudlin, etc). Throughout her case studies, Morrison sought a theoretical parameter or the "machinery" that unified. A good example of mathematical unity is the one achieved through the tensor calculus (used heavily by Kaluza, too) "which allows us to represent the unity of the electric and magnetic field" by expressing the transformation of the $F_{\mu\nu}$ tensor (Morrison 2000 191). Mathematical unity, we are told several times in Morrison's book, is not enough to warrant unification. In commenting on the unity of electric and magnetic fields within a Minkowski spacetime, Morrison echoes one of Feynman's worry that quantitative derivations do not bring understanding of the "machinery" of unification. She wondered whether the unification of electric and magnetic field in $F_{\mu\nu}$ achieved by the covariant formulation of **EM** were really "as powerful as we are led to believe" and that we needed "[…] great deal more machinery than just the transformation equations to complete the picture; there are also non-trivial empirical assumptions lurking in the background that make mathematical unity seem less grand that we might think" (Morrison 2000 190). Moreover, her follow-up question, important for my analysis of the relation between unification and explanation, is "whether or not the mathematics actually functions in an explanatory way with respect to

[213] For example, R. Batterman showed that in the case of optics, for some special optical phenomena, the reduction to electromagnetism is incomplete and an intermediate theory emerges when we take wave theory and the ray optics at the limit (Batterman 2002 ch 6). This is even a subtler matter in the case of the reduction of thermodynamics to statistical mechanics. A comprehensive discussion of this reduction by identification can be found in (Sklar 1993 337-341 and 348-373).

the kinematics/dynamics of the **S[T]R**." She concludes that in Minkowski's approach we have a formal structure and a particular parameter playing a unificatory role, although space and time, electric and magnetic fields remain physically distinct, but "united in a mathematical framework that integrates them in a seamless way" (Morrison 2000 191). In Newton's case, unification of all phenomena is the presence of one and the same force that pulls together all these bodies, a common cause. In Minkowski, we are looking for something common that unites electricity and magnetism. The Lorentz transformations explain how to relate different frames, but they do not explain "the way systems are constituted" (Morrison 2000 191). The question Morisson left unanswered is whether there is a privileged sense in which such entities exist independently of one another.[214] They provide a *synthesis* of space and time, or of electricity and magnetism, and not a reduction. If space and time are reduced to spacetime is an interesting and challenging question that can be asked in the context of Kaluza and Klein. Minkowski's theory does not provide an integration of physics and geometry, as Einstein's **GR** does. We can generalize this result now. Let us put things together now. Minkowski suggested that there are several descriptions of the world: the 3-D and the 4-D descriptions:

[67] *The world admits a 3-D representation.*

[68] *The world admits a 4-D representation.*

[69] *The world admits a* ***5-D*** *representation*

and so forth…

[214] Thanks to Christian Wüthrich for making me aware of this problem.

This is accepted by philosophers and physicists. If one is not bothered by fundamentalism here, one can ask questions about the adequacy of these representations (Petkov 2007 116):

[70] *Is [68] more adequate than [67]?*

[71] *Is [69] more adequate than [68]?*

The majority of philosophers think that [70] is not context dependent and that the answer to it is definitively yes: [68] is the most adequate description. A possible exception to this claim is some programs in quantum gravity in which the 3D+1 representation is more adequate than the 4-D representation.[215] But this is an exception more than the rule. In this thesis I tackle [69] when compared to [68]. Another case of two equivalent theories:

[72] *The classical mechanics of a system of N particles can be described by the Lagrangian formalism (with a dimension of 6N) by one equation.*

[73] *The classical mechanics can be described by the Newtonian formalism (in 3-D space) by N equations*

Here the question is:

[74] *Is [72] more adequate than [73]?*

Both attitudes are radically different and they reflect the centrality that **4-D** plays in the development of science, compared to the advancement given by the Lagrange formalism. This question can be answered only for a given domain or problems and it has few foundational consequences, although it may be important for the practice of science. In general it is meaningless and absolutely context dependent. Of course, computational or

[215] See (Wüthrich 2006 Ch. 4) for details.

representational concerns can sway the answer toward [72] or toward [73].[216] A careful analysis would reveal again that the distinction between the dimensionality of our description of the world and the dimensionality of the world in itself can be treated separately.

Where do we stand with the Kaluza's theory? Are we in case of [70] or in the case of [74]? It is important to clarify is what context [71] has an answer in the positive. If we have a world with gravitation and electromagnetism only and if we do not quantize gravity and electromagnetism and if we are interested in a vacuum theory, then the answer to [71] is affirmative. Witness the conditional nature of this answer: in general, there are no reasons to accept [69] over [68]. The arguments for **4-D** in metaphysics are also strong enough to be widely accepted.[217] In order to provide an answer to this question we need to take a look at Kaluza's unification procedure.

Christoffel symbols. I already presented the main ingredients of Kaluza's unification. His argument is that unification of **EM** and **GR** is possible in **5-D**. How? Remember assumption [25] (p. 224). In my interpretation of Kaluza's unification, the core of Kaluza's unification is the identification of Christoffel symbols with the electromagnetic $F_{\mu\nu}$. He realized that some "sectors" of the g_{mn} tensor transform like **EM**-tensors.[218] After postulating [24] and as a result of **(71)**, Kaluza suggested that $F_{\mu\nu}$ is a "degenerate" (*verstümmelte*) form of the **5-D** Christoffel symbols and proceeded to identify them by the identifications of **EM** tensors and vectors with parts of the g_{mn} in . The electromagnetic

[216] As I mentioned, J. North questions in a recent unpublished paper this equivalence.
[217] See for example (Sider 2001, 255) as the definite reference to a four-dimensionalism in metaphysics.
[218] In this analysis I prefer to use the term "sectors" for various components of g_{mn}. Although part is a legitimate concept, I prefer to avoid the confusion with the technical concept of part in the mechanistic philosophy.

field strength $F_{\mu\nu}$ enters the Γ^{λ}_{rs} when only one of r or s is 4, whereas the theory cannot specify what kind of symbol can be employed in the last term. It has to be a scalar field that enters the $\Gamma^{\lambda}_{4\mu}$ only when one of the λ or μ is 4.

Remember that Kaluza operated with Christoffel symbols and not within an action principle as did Hilbert. Christoffel symbols are first order derivatives of the g_{mn} tensor and they are related to the curvature. He intuited that the Christoffel symbols in 5D transform like curls and the can be identified with the $F_{\mu\nu}$ according to $\mathbf{ID_1}$. Kaluza intended to preserve the 0-3 part of the Christoffel symbols and to give $\Gamma^{4}_{\mu\nu}$ an "electrical" interpretation.

Kaluza's crucial step forward is that he interprets geometrically $F_{\mu\nu}$ –as a special type of geometrical connection. As a consequence, $g_{\nu 4}$ look like the A_{μ}. The nice, clean modularization of g_{mn} is a result of the identifications of $\mathbf{ID_1}$. In Kaluza what is basic is the form of the Christoffel symbols, i.e. the features of the Levi-Civitta connection. There is neat preference for geometry in Kaluza's approach over physical meaning. Some authors ignore this detail based on the alternative formulation based on the Lagrangian and action principle (O'Raifeartaigh 1997 48). The fact is that Kaluza did not start from the Lagrangian, nor from the transformations as Klein did, on the contrary to what is usually thought.

Once the details of $\mathbf{ID_1}$ are worked out, the 4×4 part of $g^{(5)}_{mn}$ can be identified with $g_{\mu\nu}$. So, where is $F_{\mu\nu}$ to be placed? The simplest way is to divide in three *sectors* as follows:

$$g^{(5)}_{mn} = \begin{pmatrix} g_{\mu\nu} = \text{G sector} & g_{4\nu} = \text{EM sector} \\ g_{\nu 4} = \text{EM sector} & g_{44} = \phi = ? \end{pmatrix} \tag{139}$$

which can accommodate the $g_{\mu\nu}$ tensor in the 'G' sector as well as the A_μ vector in the 'EM' sector.

What he did was similar to a division into functions: this sector here has this role; the other sector there has that other role; the last sector of ϕ will remain un-interpreted. I call this procedure the *modularization*.

This intuition of Kaluza goes beyond simple mathematical identifications. The theory of gravity was based on the local coordinate invariance. But **EM** was based on a local *internal* symmetry, the gauge symmetry. There is something different about the **EM** transformation. What Kaluza did not know was that not only **EM**, but other interactions are "gauge" theories with local symmetries. What Kaluza intuited, Klein expressed much clearer five years later: once we add dimensions to the spacetime manifold, the concept of local internal symmetry of electromagnetism can be derived from a local coordinate invariance of the **5-D** manifold. One can pack Kaluza's theory in the language of "gauge invariance" but even so the condition **CYL** still look unphysical.

The suspicion that Kaluza's theory is closer to the conjunction is the form of the metric (13). He managed to represent within a **5-D** tensor both electromagnetic and gravitational interaction without strongly *mixing* them.[219] How close is Kaluza of a trivial paste structure such as ? Think of the covariant form of **EM**: in the form of the electromagnetic stress-energy tensor **(13)**, there is a difference between electric and magnetic fields. The electric field resides only on the columns are rows corresponding to the zeroth line, that corresponding to time. The magnetic field fills the part of the metric that corresponds to spatial dimensions only. One can see why the amalgamation of electric and

[219] A concept used in that period was "coalescing" or "amalgamation" of **EM** and **GR**.

magnetic fields is not total in the case of the covariant **EM**. Maxwell tensor discriminates two sectors: the electric and magnetic sectors. Similarly, in Kaluza the electromagnetic field depicted by A_μ resides only on the rows and columns corresponding to the fifth dimension. This means that the fifth dimension is the source of the appearance we see in 4-D of the electromagnetic world.

Several sectors of Kaluza's g_{mn} may react differently to coordinate transformations. Some parts of it can be brought to a familiar form, i.e. the type of transformation a stress-energy tensor display to a given coordinate transformation. This is based on a deeper assumption according to which the way a specific tensor transform is essentially linked to the type of the theory it belongs. **EM** is then characterized by the way $F_{\mu\nu}$ transforms under the coordinate transformations permitted: $F_{\mu\nu}$ is the tensor of electromagnetism and transform accordingly. What is the underlying assumption here? If a given tensor transform like the **EM** tensor it can be identified with the EM field. Within the **5-D** g_{mn} one can see how its sectors components are *like* electromagnetic tensors. In other words, if a sector transform of a larger theory T like a known tensor belonging to a theory T_1, theory T_1 is now part of the larger theory T. Is this reduction? In fact it is not because T_1 has its own conditions to be imposed upon T and in this case it is related to the interpretation of p^4 for example. Without T_1, p^4 would not be interpreted at all.

Similar to Maxwell's case, one can see why we have the illusion of **EM** and **GR** as disparate theories: once cylindicity **(CYL)** is assumed, Kaluza represented the **EM** and **GR** interactions under one and the same formalism and inferred a geodesic equation; from **ID$_1$** he inferred the form of the metric tensor g_{mn} and from **ID$_2$**, the geodesic equation for macroscopic objects; **ID$_3$** had helped him to provide an interpretation for p^4. The above **ID**s provide answers to "why" questions such as: Why is it apparent that **EM** phenomena are independent of gravitational phenomena? Why do macroscopic charged particles not move on geodesics in 4-D? Why do **GR** and **EM** obey Poisson equations?

In Chapter 16 I address a different question: do we gain anything in understanding the world by answering such questions, by analyzing Kaluza's explanation when compared to Klein's. This immediately raises philosophical questions about what it means to *observe* a direction in the spacetime manifold, but let us postpone this until later. Let us accept the commonplace claim that we experience three dimensions of space and one of time. Small or null variations of the field on the fifth dimension means that the world is "cylindrical" on the fifth direction in the sense that every point $P(x^0 \ldots x^4)$ can be identified with another point P' having the coordinates $P(x^0 \ldots x^4 + dx^4)$ if *all* fields and *all* derivatives are smooth on the fifth direction (this analogy was proposed in (Einstein and Bergmann 1938, 683). Their identification is not a absolute identification, it is only a way to explain the non-observability of the fifth dimension. Points P and P' are still distinct, but the values of the *g* field are equal or have close values at these points. Let us think that objects in this classical field theory are formed by the value of the field *g*. Then field objects fill the whole x^4 space. They are elongated objects that never end in x^4.

In answering [65], we want to know what the symmetry of Kaluza's **5-D** manifold is. By assuming that physical fields do not change on the fifth axis, the topology of the fifth direction is not affected; it is still the standard topology of real numbers and not that of a circle and the homeomorphism with $\mathbb{R}^5$. By this, the "external" symmetries of spacetime are the same with those of **GR+EM** and the identity of spacetime points is not affected and the [27]. This is a major difference to Klein's **COMP** condition and this is why we cannot talk about compactification in Kaluza.

The empirical condition imposed is that the fields in the manifold are such that there is no difference between their values along the x^4 axis, assumption which does not affect the identity of points on it (Duff 1995 3).

13.4. Does Kaluza's theory achieve unification?

There may be several ways of arguing that Kaluza's theory is trivially similar to **GR**. Kaluza kept the Riemannian metric in **5-D** which suggests that Kaluza's theory is nothing more than gravitation in **5-D**. In the same spirit, one may ask whether Kaluza achieves more than a mathematical unity and whether it provides the unification linked to explanation Morrison was looking for. If one buys literally Morrison's criticism, there is no theoretical parameter in Kaluza to perform the unification, no "real unificatory element" or "machinery" (such as the "displacement current" in Maxwell). Kaluza depicts a mathematical operation that unifies and one cannot identify a causal agent that unifies. By using the cylinder condition (**CYL**) and specific approximations, Kaluza showed that parts of a five dimensional manifold can be identified with the 4-D gravitation and parts of it with the electromagnetic fields. But there is something more than a mathematical trickery in Kaluza. He used a form of modularization of the metric tensor that will become an inspiration for later work in particle physics. One can find modularity in mechanistic philosophy. In the original work, Kaluza did not start from the metric, but from the form of the Christoffel symbols which are roughly speaking second derivatives of the metric tensors and have a geometrical interpretation of parallel transport. Kaluza extended gravity from D=3 to D=4, but this time there are two theories involved in the process and he replaced a 4-D gravity theory and a 4-D **EM** theory with a **5-D** theory of a field in which physical fields are "smooth" on the fifth dimension. Last but not least, Kaluza's project was to unify two types of physical interactions; Minkowski unified space and time. We cannot speak properly of a modularization in Minkowski, but Kaluza's separation of sectors is clearly the result of idealization and simplification which can be judged as *ad-hoc*. If some similarities between Minkowski and Kaluza are evident, there are major differences, especially in the context of unification. What is indeed encouraging is the possibility of inferring the laws of **EM** directly from the geometrical form of the Christoffel symbols in **5-D**. The field theory in **5-D** is not a mere gravitation, but a gravitation restricted by **CYL** and in fact is not gravitation anymore.

There is also another argument that ties **GR** as we know it to the **4-D** manifold. Although we can go down in conceiving gravitation in 3-D (see Section 15.8), we have reasons to believe that it is profoundly unnatural. The so-called extension to other dimensions of our 4-D **GR** is also less realistic. The mathematician would think that physical theories are like geometrical objects that can be generalized from D=3 to any D. This is wrong and sets specific constraints on the naïve geometrization program. I claim that Kaluza did not adopt this naïve extension. He saw that there is no such a thing as a naïve **5-D GR**. The 4-D **GR** came scathed from Kaluza's unification. Here are at least three reasons to see the **5-D** field as being different of Einstein's theory and not as a mere conjunction of **EM** and **GR**:

> ***[75]*** *The interpretation of u^4 is not similar to u^0, u^1, u^2 and u^3. Kaluza's theory is open to interpretations not available to the 4-D **GR**.*

[76] *The presence of the field ϕ;*

[77] *The different kinds of symmetries **EM** and **GR** have. Given **CYL**, the **5-D** metric have terms not present in the four-dimensional **GR**.*

In regard of [75], I emphasize that in Kaluza there is a new interpretation for the velocity on the fifth axis. A particle does not change its p^4 momentum in time as it is proportional to its charge. If charge is an invariant of the motion of a free particle and it does not depend on its energy, then u^4 is constant in spacetime. This is in contrast to u^0 and all u^1-u^3. Kaluza's theory in fact is open to the possibility of a variation of u^4 in spacetime, although once q is fixed by other means—mainly by a postulate of the conservation of charge—the fifth component of u is constant.

The x^4 direction is special or "different". Instead of providing a direct answer, I want to mention again that adding time to the three spatial dimension is not a spatialization of time. This fact has an interesting history in itself. Meyerson asked Einstein during a meeting of the French Philosophical Society in 1922 whether the spatialization of time is the right interpretation of **SR** (Meyerson 1985 (1925), 268; Nahin 1993 264). According to Meyerson, Einstein answered tersely: "it is certain that in the four-dimensional continuum all dimensions are **not** equivalent". What is at stake is not Einstein's position on the tense-tenseless debate. If one believes that time is not the same as other spatial dimensions, then it might be the case that the standard block universe model is not the right one. Einstein is usually quoted by saying the physics is "a happening" in 3-D+becoming and "an existence" in 4-D (Einstein 1961 p.).[220]

[220] Arguably, Einstein adopted later on in his life the block universe view.

In the case of Kaluza's extension to **5-D**, if x^4 has a special interpretation then one can suspect that the arguments purporting to show that we live in a D=3 spatial dimensions are correct by adding a proviso: "we live in a universe that has exact D=3 of *this* type of spatial dimensions". Such a qualification is necessary for the transition from 3-D to 4-D. As one can easily realize, [76] is another troublemaker for the Kaluza-Klein theory, but I interpret it here as an element that signals that unification is neither reductive, nor trivial. Think again by analogy: new theories frequently use new parameters that are left un-interpreted. **GR** has for a while the infamous constant Λ which has an interesting history in itself. Even now it is not clear whether it should be set to zero and disbarred forever or on the contrary to look deeper for its significance and whether it tells us that there is something wrong with **GR**. The field ϕ signals that the unification adds something to the conjunction of **GR** and **EM**, although it is not at all a desirable new element. Another new element of Kaluza's theory is the fifth dimension, or weaker, "world parameter".

Can we dispense with the fifth dimension? Let us take a look at the static universe in 4-D, i.e. a universe without time dimension. Christoffel symbols are:

$$-2\Gamma_{\mu\nu 0} = \partial_\mu g_{\nu 0} + \partial_\nu g_{0\mu} - \partial_0 g_{\mu\nu}$$

Kaluza's unification *cannot* be mocked in 4-D. If one sets $\partial_0 g_{\mu\nu} = 0$, one can recover a "static electromagnetism" in 4-D in which $\partial_0 A_\mu = 0$. As there is no such a thing as "static" electromagnetism and static gravity in 4-D cannot incorporate electromagnetism, and the **5-D** gravity stationary on the fifth dimension, i.e. satisfying the **CYL** condition is able to do so one can see why there is less triviality in Kaluza than one might suspect at the first sight.

The dynamics of the ϕ field is not trivial at all and the analogy with static gravitation in Kaluza is in fact possible and it was subsequently discussed. Kaluza left this field un-interpreted but later it was called the “dilaton field”. Up to now, it is a hypothetical field without an observable particle and discussing it again in the context of Klein’s theory would be a better option.

In respect of [77], there are several things to say, but I prefer to defer this discussion to the next chapter where I’ll discuss the problem of symmetry reduction in Klein.

I conclude that there is unification in Kaluza’s theory beyond a mere conjunction and there is no reduction of **EM** to **GR**. One can still think that that there is reduction of **4-D EM** to a **5-D GR**, but I showed that the **5-D GR** is not a mere generalization of the **4-D GR**. Despite this facts, *compared to Klein,* Kaluza is relatively close to what a trivial unification is, but I argue that Kaluza’s theory cannot be assimilated to a mere conjunction and ϕ even left uninterpreted plays a crucial role. All in all, ϕ is an *ad-hoc* element, as is the **CYL** condition. Kaluza added other ad-hoc elements in trying to recover trajectories of both macroscopic objects and elementary particles.

There are the most important ingredients of unification in Kaluza, which is far from a perfect unification. As I argued, it is neither a mere conjunction. What is Kaluza’s result? He was able to infer the trajectories of charged particles from geodesics in **5-D**. I take it as the major consequence of Kaluza’s unification. It is a result of unification and not properly speaking an explanation. In respect of [56] and [57]: are there novel explanations and predictions in Kaluza? This is a question to be addressed later, but for the time being I claim that we deal only with results and consequences of unification in Kaluza and not with genuine explanations. In the following chapter I place Kaluza's and Klein's theories in Maudlin's scheme once I will discuss in details the symmetries associated to both of them.

Perhaps the general lesson Kaluza teaches us is that the trivialization by conjunction is a matter of degree, and that the very concept of dimensionality can be rethought in terms of unification. What are the other virtues of Kaluza's theory besides parsimony and economy of thought?

Chapter 14. Unification in Klein

Kaluza tried to corral unification to mathematical operations on Christoffel symbols, although he did not realize all the inherent problems of interpreting physical measures on x^4. In this chapter I focus on the difference between Klein and Kaluza by emphasizing the major improvements of the former. I argue here that Klein's theory is more unificatory than Kaluza's.

Although Klein applied different procedures to unify **EM** and **GR**, this is not the main result I want to focus on: Klein came with a new interpretation to what is important in adding x^4; he showed how to use the nascent formalism of quantum mechanics in the 5-d manifold; he explained several brute facts of **EM**.

Klein was slightly influenced by Kaluza. We know he read Kaluza's paper only after he had elaborated the theory. His spirit of resignation after reading Kaluza's article is only a sign of his own modesty (see Section 11.1 for details). In this chapter I show in what sense Klein's unification is (a) different and (b) stronger than Kaluza's because he solved many of the difficulties of the previous approach and improved it significantly. Ranking unification from mere conjunctions to "perfect unifications" in Maudlin's words, reveals a variety of instances in which unification is neither perfect, nor trivial. Both Kaluza and Klein fall within this rich variety of unificatory theories.

There are several reasons to emphasize the novelty of Klein's work. First, the hope for a unification between **QM** and the theory of fields (**GR** and **EM**) (if such a thing exists, it can be anointed "*the* greatest unification ever"), as one of the most prominent conceptual problems in the philosophy of physics took a clear form in Klein, even in the wake of the quantum program itself (Schrödinger's paper on quantum mechanics was published in 1925). I argue that Klein's main motivation was to connect the five-dimensional formalism to quantum physics. Remember that the Klein-Gordon equation was formulated in the 1926 paper as a covariant equation of a quantum particle. Then, because of the compactification condition (**COMP**), the masses of photon and graviton were suggested as mean values of fields, a step toward the relation between theory of matter and theory of fields. As some authors suggested, Klein's 1926 article can be considered a precursor of non-Abelian gauge fields (O'Raifeartaigh and Straumann 2000, 1-23; O'Raifeartaigh 1997). String Theory, as it is formulated nowadays, relies on the procedure of "compactification" proposed by Klein.[221] Last but not least, some string theorists pay a certain tribute to Kaluza-Klein theory, whereas few philosophers discussed their model of unification.[222] Although this is not the theme of the present chapter, we have to keep in mind that strictly speaking, even in its most liberal reading possible, Klein's theory is false. I argue that it has several resources which were explored by few historians and philosophers.

Kaluza and Klein are depicted in several popularization books as the first physicists having the idea of compactified dimensions (Randall 2005 ch. 2; Greene 1999 ch. 8). This is false: there were other authors who speculated on this issue (C.S. Hinton, G. Nordström *i.a.*) and strictly speaking Kaluza did not talk about compactification. As I argue here, Klein is the father of the compactification procedure itself—both historically and con-

---

[221] In (Klein 1928) he came back to the problem of the unification and restated the main idea of compactification in direct relation to conservation laws.

[222] Some standard textbooks in String Theory extensively discuss Kaluza-Klein. See (esp.Ortìn Tomás 2004 ch. 11; Polchinski 2005 vol. I, sect. 8.2; Kiritsis 2007 app. E). L. Susskind discusses the Kaluza and Klein (although he makes the confusion between Kaluza and Klein by claiming that "Electrical charge in Klauza's theory is quantized" (Susskind 2005 235) in a chapter called "Reincarnation" with a direct reference to the resurgence of Kaluza-Klein theory.

ceptually. Beside such inconsistencies, it is clear that for the vast majority of string theorists, both Kaluza and Klein are the harbingers of String Theory. I take it as a major source in the analysis of unification and a mine of philosophical issues regarding the relation between unification and explanation.

In Part II, I suggested some differences between Kaluza and Klein and I decried the superficiality of the literature on this matter—exceptions are notable though. To be fair, there are some similarities between Kaluza's and Klein's theories, especially at the formal level: the main structure of Klein's argument is the same as Kaluza's. He tried to unify **EM** and **GR** in the vein as Kaluza. Klein's geometrization is also similar to Kaluza and he was conservative in respect to pseudo-Riemannian metric (41) of the manifold and its signature. But in order to argue for Kaluza's supposition that the fifteen quantities of the symmetric tensor g_{mn} would accommodate the ten independent components of $g_{\mu\nu}$ plus the four components of A_μ, Klein started from a different perspective, i.e. the coordinate transformation in 5-D.

As I see it, Klein's unification is the result of several factors some of them being innovative compared to the existing attempts. One witnesses in Klein **COMP**, the change in the topology of the fifth axis from linear to circular topology and this has a major impact in the unificatory power of Klein's theory. The consequences on the unificatory power are:

[78] *The role of the coordinate transformations and their invariance;*

[79] *The factorization of the action (i.e. the action is written as a product of two independent terms);*

[80] *The behavior of the wavefunction in 5-D;*

[81] *The symmetry of* ***EM*** *becomes a symmetry of spacetime;*

[82] *The **stronger EM-GR** coupling within* g_{mn}.

None of which were present in Kaluza. Perhaps there are some other reasons to conclude that Klein's theory is not a notational or procedural variant to Kaluza's. My argument in the following chapter is that in the unification context, Klein's results, consequences and explanations, are dramatically improved. Klein's explanatory store is richer than Kaluza's and his theory does not impose strong approximations anymore, as Kaluza's did. Klein also took off the slow motion constraint and substantially relaxes the condition of weak fields. All these had a major impact on the explanatory power and the unificatory strength of his formalism.

14.1. Invariance and factorization of the 5-D action

Remember that Kaluza put an emphasis on the metric and the forms of the Christoffel symbols. Klein's approach was more innovative than this as he used [78]; start from the invariances of the coordinate transformations and hypothesize the metric. He got closer to the spirit of the gauge transformation discussed by Weyl. There is also a philosophical aspect of this procedure. Hilbert was adamant about the emancipation of formalism from its dependence of coordinate transformations. In one of his lectures, he stated that "a sentence about nature, expressed in coordinates, is only then a proposition about the objects in nature, if the sentence has a content which is independent of the coordinates" (cited in Majer and Sauer 2005 270). By **CYL**, Klein showed how this emancipation could be worked out, at least partially by employing the action, which is invariant to coordinate transformations. Hilbert suggested that coordinates were necessary elements of the way in which represent the world, but that physical theories should look for coordinate free formulation, whenever they are possible. Recall that Kaluza-Klein theories set aside the covariance because the fundamental assumptions needed are not covariant and this can be interpreted as a major step back to a previous stage.

Why is invariance of the coordinates still important? Think of "stationarity". **SR** teaches us that there is no such a thing as stationary objects. In a different frame, stationary objects are moving and *vice versa*. But still, in **GR** we use stationary solutions to the cosmological problems. The same can be said about the fifth spatial coordinate. If the first four coordinates are identified with observable coordinates, then for empirical reasons we want to minimize the influence of the fifth coordinate which is not observable. In other words, can we make the universe x^4 "stationary", in the spirit of the discussion of "stationary universes"?[223] In the stationary models we assume that the world is frozen in respect of time. Can we freeze the fifth coordinate? By this, one can have a different perspective on the **(CYL)** condition in Kaluza. In a group parlance, one can say that the subgroup of the transformation of the x^4 axis leaves invariant the components on the axes $x^0 \ldots x^3$. One way to do this is to restrict the class of possible transformations to those in which $x^0 \ldots x^3$ do not transform as functions of x^4 but only as functions of $x^0 \ldots x^3$. This means that all vectors $V^\mu(x)$ of a spacetime surface $x^4 = ct$ have two parts:

- A spatial-temporal part that is invariant to any transformation of the x^4 axis but depends on the symmetries or the transformations of coordinates of the $x^0 \ldots x^3$ axes.

[223] Here x^4 is not a timelike coordinate and one can balk at the usage of stationarity here. I use the term "stationary" in a strict sense of derivative on the fifth direction: ∂_4 similar to its usage for the time axis: ∂_0

- A 5-D component that transform as a function of $x^0 \ldots x^3$ and as a function of x^4.

I take [79] as part of Klein's unification attempt, although it is a standard procedure that can be found in any part of physics: when one of the coordinate is periodical, the action can be separated in several independent components each part corresponding to a dimension of space. It is not the action principle that makes Klein interesting here, but the fact that he used factorization in the analysis of **EM** field. Factorization of the action is a common procedure when for example the **EM** field in a box with reflective walls or when the hydrogen atom is represented in a system of polar coordinates. Periodicity or boundary conditions entail factorization. In itself, it is a computational procedure: it is not clear whether it says something about the world or the way in which some fields act differently in respect of some coordinates.

Klein used the action principle as discussed in the context of Hilbert's approach. The principle has its own heuristic importance and it could be applied to any theory, including Kaluza's. In Klein it is important to see how the action *factorizes* in terms which reflect the action in **4-D** and one constant action in **5-D**. In other words, action is split into two terms. If one thinks in terms of breaking the symmetry, this is the place where the nice symmetry $\mathcal{M}^5$ of **5-D** is broken. I mentioned in Section 8.4 the whole debate surrounding Hilbert's theory in the case of the **4-D** gravitation and whether Hilbert's procedure is equivalent to Einstein's. The action can be written as a sum of two or three integrals (up to some constant factors). The first two integrals are:

$$\begin{cases} S_1 = -\int d^{(4)}x\sqrt{-g}\ R \\ S_2 = \int d^4x\sqrt{-g}\,\frac{1}{4}\phi F_{\mu\nu}F^{\mu\nu} \end{cases} \tag{139}$$

S₁ is simply the action for gravity in 4-D, while S_2 is an action of the electromagnetic field of a stress-energy tensor given by Maxwell equations and S_3 is the action of a Klein-Gordon type of particle having the scalar field ϕ (O'Raifeartaigh and Straumann 2000 9; Overduin and Wesson 1997 15).

When the field ϕ is not set to zero, an extra term has to be added to the action:

$$S_3 = \int d^4x\sqrt{-g}\,\frac{\partial^\mu\phi\,\partial_\mu\phi}{\phi^2} \tag{139}$$

Once Klein ascertained to (105) and (108), the integral in (97) factorizes in the sense that it is the product of two integrals $\int dx^{(4)}\int R^{(5)}\sqrt{-|\hat{g}^{(5)}|}d^4x$ with the first integral being a constant.[224] In Klein's theory, S_3 is a constant so it never contribute to the variation of the action such that Klein's action has only two terms. By minimizing the action, $\delta S_{Klein} = 0$, the result is a system of two equations (Klein 1926 898):

$$\begin{cases} R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R = -\frac{\beta^2}{2}T^{\mu\nu} \\ \partial_m\sqrt{-|g|}F^{\mu m} = 0 \end{cases} \tag{139}$$

where $T^{\mu\nu}$ is the contravariant component of electromagnetic energy-momentum tensor.

In order to obtain the Einstein's field equation, one has to set:

$$\beta^2 = 2\kappa ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu - \left(\frac{R_4}{\lambda_4}\right)^2 dz^2 \tag{139}$$

---

[224] I use some of the materials from the site of the 1998 conference on Klein: *Oskar Klein Meeting: D>4 , t=75*, (Ann Arbor, 1998): http://feynman.physics.lsa.umich.edu/klein/newklein.html

The argument based on action had several virtues compared to the procedure used by Kaluza. Klein's machinery dramatically simplifies the computation in **5-D** which would be difficult to guess without the "action principle". It can be said that the symmetries of spacetime can do extra work in the context of action principle when one dimension is compactified. Without factorization of the action, Klein's argument would look much more convoluted. I identify this factorization of action as another element of Klein's unification which is definitively part of its unificatory power. The new **5-D** theory nicely splits into two actions, the Hilbert-Einstein action S_1 of the **4-D** gravitation and the action of **EM** S_2.

14.2. The wavefunction in 5-D as an external element of unification

Klein employed identifications as mathematical procedures, like Kaluza did, but he went beyond this. I claim that there are two aspects specific to Klein's unification: the wavefunction as an extrinsic element of unification and the reduction of types of symmetries of the theory. While the former illustrates the theoretical entity that Morrison demands for unification, the latter is connected to Kitcher's perspective on unification. Both are, I argue, crucial to understanding Klein's improvements upon Kaluza as Klein's new argument and the unification he achieved were more powerful than Kaluza's.

The wavefunction is an extrinsic element in this case because Klein's conviction was that his 5-D theory can say something about the dynamics of real particles. He moved toward a "theory of matter" based on wave mechanics and this is in sharp contrast with Kaluza's "vacuum theory". According to de Broglie, real particles are described by the associated "matter waves". But the route to matter from vacuum was a real toil. We saw that Klein tried to infer the dynamics of charged particles from the "ray" of the wave function in 5-D. This is of course not present at all in Kaluza: after all, de Broglie's thesis was published in 1924 (de Broglie 1924, 111, [1]). What is important is that Klein related the electrical charge to the fifth dimension. Other authors relate the extra dimensions to other material features of elementary particles. F. London for example associated the spin of the electron to the fifth dimension. Eddington also ventured into deriving the Schrödinger equation from the **5-D** theory.[225] But all these results were dependent upon Klein's idea to write Schrödinger equation in **5-D**.

In the previous chapter I showed that Kaluza's theory is not a mere conjunction of theories. In my interpretation, the central element of unification in Klein is the behavior of the wavefunction in 5-D which is an extrinsic element to both **GR** and **EM**. The wave-function in itself is not part of these theories, so Klein's unification cannot be deemed as a mere conjunction of **GR** and **EM** either. It plays the role of the displacement current in Maxwell and it is associated to a mathematical structure, i.e. the Sommerfeld condition of stationary on a closed orbit. This mathematical condition plays afterwards a heuristic role in the discovery of compactification which, as a topological condition, is compatible with both **GR** and **EM**. I want to stress that the wavefunction in 5-D, undoubtedly inspired by de Broglie's *Ansatz*, is *not* an electromagnetic wave or a gravitational wave *per se*. It is arguably far for being Schrödinger's wave function either. Being central in the new argument, **COMP** is a unificatory structure equipped with explanatory powers. It comes from wave mechanics or, from a modern perspective, from the formalism of quantum mechanics

[225] For an excellent historical approach to the quantum mechanics in 5-D after Klein, see (Goenner 2004 sect. 7.2.4) Goenner does not insist on Klein's development but provides excellent historical analyses of later developments: Fock, Mandel, Dirac, Einstein, etc.

in de Broglie's interpretation. One may ask whether this extrinsic element of unification is specific only to Klein's unification. It is worth knowing in general whether the element that generates the unificatory theory T is intrinsic to T_1 or to T_2. Klein demonstrates better than any of Morrison's examples the importance of the "extrinsic" element of unification, i.e. an element that unifies but is not part of T_1 or T_2.[226]

14.3. Symmetries in Klein's theory

The other consequence of **COMP**, [81], is an important result in Klein which is illuminating in respect of both unification and explanation. I discuss this aspect here in the context of unification and revisit it in the context of explanation. (see 16.7). In the 1920s-1950s, physics operated based on two fundamental ideas of invariants: one is the local coordinate invariance discussed in **GR** and in Klein; the other is the local internal symmetry illustrated by the **EM** and later on by the gauge theories of strong, weak interactions: virtually all theories of physical interactions are gauge theories. Even before the discovery of gauge aspects of physical interactions, Herman Weyl claimed that "As far as I can see, all *a priori* statements in physics have their origin in symmetry" (Weyl 1952 126). If we find the maximal specifications of the internal symmetry of the theory, we can characterize theories in terms of their maximal set of internal symmetries.

Coordinate transformations of space and time are called external symmetries. They do not belong properly speaking to a theory, but to the representation of space and time and to the representation of an object in space and time. Internal symmetries belong to theories. External symmetries are properties of objects or spacetimes but not of theories. In quantum

---

[226] Maybe another extrinsic element of unification is the "string" and the "brane" in string theory that are extrinsic elements to both the standard model and to the theory of gravity. I do not claim that an "extrinsic element" of unification characterizes *any* unification.

mechanics, a permutation of particles is an internal symmetry. Ditto the change of A_μ in covariant **EM**. Even more intuitive is the invariance of laws of electrostatics to the absolute, global change of the electric potential. The whole theory depends on differences in potential, not on absolute values (if such a thing exists). One important point is in order here: this distinction between external and internal can be problematic in the case of **GR**. Wigner for example thought that the symmetries of **GR** are dynamical, i.e. they belong to the theory, not to space and time (Wigner 1967, 280). The received view now is to take these invariants as external (Earman 2004, 1227). Without going further into the details of this discussion, it is worth noting that according to the economy of thought professed by Mach and akin to the unification present in Klein, we should represent the world based on external symmetries or at least to prefer external them when internal symmetries cannot be eliminated.

Sometimes the active/passive distinction among symmetries is considered relevant only in the context of coordinate invariance of the theories. A coordinate transformation is *passive* whereas the transformation of the manifold itself is *active*. A coordinate transformation in a generally covariant theory such as **GR** is another representation of the same system. Performing the diffeomorphism transformation on the manifold renders a different system, i.e. it is an active transformation.[227]

Last but not least, symmetries can be local or global. The local symmetries are more general because in their case the transformation is a function of space and time. Lorentz transformation is global, whereas all gauge symmetries are local (Kosso 2000, 81-98). Newton's laws are invariant to some specific transformations: rotations, transla-

[227] This originated the well known "hole argument".

tions, translation in time, time reversal, etc. These are all global transformations because invariance is obtained when we apply the same transformation to all points in space and time. The covariance of **GR** is roughly its feature to preserve the form of the equations under the action of a transformation group. It is related but not identical to the invariance group of **GR**: there are theories with displaying covariance but having a different invariance group.

***EM** as a gauge theory.* Klein was able to show that the two types of invariance were not independent. Even before the formulation of **EM** as a gauge theory, Klein showed that Maxwell's action can be factorized out of the action in 5-D. But before discussing Klein's unification of symmetries, it is useful to see in what respect **EM**'s internal symmetry is recovered in Klein's theory as symmetry of spacetime. According to Klein, the invariance of **EM** can be inferred from the symmetry of 5-D spacetime. In this specific sense, there is a reduction of the types of symmetries the theories need.

EM dynamics is intimately related to its symmetry principle. Electrostatics is a theory that depends on the difference of the electric potential in two places, not of its absolute value. More rigorously, one can say that the theory should retain the form of its equations if a transformation of V to V′ is performed:

$$V' = V - \frac{\partial \chi}{\partial x} \qquad (140)$$

Mainly by subtracting a number from the potential, the description remains the same.[228] The main idea is that physics remains the same if we add a constant to the electric potential. The same can be said about the vector potential *A*. An important idea in the de-

[228] The partial derivative of χ is not important here.

rivation of the covariant form of **EM** was the introduction of the vector potential *A in lieu* of **E** and **B**. For two physical fields **E** and **B**, the values of *A* are not unique. Two values of A that differ only by the gradient of a field χ will give the same values of **E** and **B**:

$$A'^{\mu} = A^{\mu} - \frac{\partial \chi}{\partial x^{\mu}} \tag{141}$$

This is the source of the gauge invariance of **EM**. We saw that the tensor $F_{\mu\nu}$ is manifestly covariant in the sense that it is unchanged by a transformation of coordinates (150) such that $F'^{\mu\nu} = F^{\mu\nu}$. The reason is that F is a four-dimensional "curl" operator that is left unchanged by (150). One could say that this is a sufficient reason to drop ***A*** and *V* altogether.

Unexpectedly, the reason to retain A_{μ} and *V* comes from **QM**. The interaction between a quantum particle with mass *m* and charge *q* and an **EM** field is given by a Hamiltonian depending on **A** and V not on **E** or **B**:

$$H = \frac{1}{2m}\left(\mathbf{p} - q\mathbf{A}\right)^2 + qV \tag{142}$$

Such that **A** and V are promoted in **QM** to the status of operator. The gauge transformations (150) and (149) will affect the quantum wave function by a phase factor:

$$\psi'(x,t) = \psi(x,t)e^{iq\chi(x,t)} \tag{143}$$

Do Ψ and Ψ´ represent the same reality? Yes they do, but other quantities in which one might be interested, the so called currents are not invariants to the gauge transformation and this acts like a constraints imposed upon the formalism of **QM**.

We started from the gauge transformations of ***A*** and V and we inferred the possible phase transformation of ψ. The argument can be run in reverse: the wavefunction can be changed without leading to observable effects by changing its phase *locally* because observables are related to the amplitude of the wave function. In short, by postulating (152) one can recover the gauge transformation of electromagnetism (149) and (150). So one can see that the unification between **EM** and **QM** is possible.

In the presence of **EM** field, the wave function acquires a phase factor called the "Weyl factor" which is an abstract element that needs an interpretation. F. London in fact inferred this factor as being the integral of the A_μ around a closed curve (London 1927):

$$\psi' \to \psi e^{\frac{ie}{\hbar}\oint A_\mu dx^\mu} \tag{144}$$

On the contrary, the Weyl factor can be the momentum, as Fock assumed—inspired by de Broglie. Fock factorized the wave function into a space part and the momentum part (Fock 1926 228):

$$\psi' \to \psi(x,t)e^{2\pi i\frac{p}{\hbar}} \tag{145}$$

So one can see that the wavefunction, i.e. the solution of the Schrödinger equation, is invariant to the gauge transformation. This important formal result was inferred on several grounds by O. Klein, H. Mandel and V. Fock and by Gordon between 1925 and early 1927.[229]

Schrödinger had anticipated in 1922 de Broglie's result that Weyl's scale factor (the exponential factor that relates the lengths of a rod parallel transported from P to P':

[229] It is not clear which author had the priority. Fock worked in Leningrad and was somehow isolated from the German scientific world. He received Klein's 1926 paper with a considerable delay. London published the result in 1927, but presented it in December 1926, after the publication of Klein's paper, although he was aware of the connection between Weyl's scale factor and Schrödinger's theory (London 1927, 375-389; Fock 1926, 226-232), both are translated in (O'Raifeartaigh 1997). Historical details of the London-Schrödinger letter exchange are related by C. N. Yang are in (Kilmister 1987, 253). Fock's historical priority is discussed in (O'Raifeartaigh 1997 79sqq, 94-100).

$l_{P'} = l_P exp\int \phi_i dx^i$ for closed orbits was an integral power of some universal constant (Vizgin 1994; de Broglie 1924, 111, [1]; Schrödinger 1923, 13-23). Their results are fundamentally correct, although they do not reflect the relativistic motion of a real electron. In order to understand the connection between matter and the fields, a further assumption regarding the spin of a particle was necessary. Klein did not take the spin into consideration, but Dirac did. In 1928, Dirac was able to infer the correct equation in which the spin one-half was added. This is the equation that indeed represented the real particle electron, unlike Klein-Gordon which is a spinless equation of a charged particle.

As we saw, Klein associated the Weyl factor to x^4, the fifth axis. One can see why x^4, as a phase factor of the wavefunction, is the unificatory element here. But it is associated to gauge potentials which are shaky quantities.

There is a rich philosophical debate around the reality of gauge potentials (Healey 2007, 297; Healey 2001, 432-455). M. Redhead called the gauge principle "the most pressing problem in current philosophy of physics" (Redhead 2003). The "No New **EM** properties" view is the minimalist attitude according to which there are no intrinsic properties besides **E** and **B**. The "New Localized EM properties" view adds localized **EM** properties on the top of **E** and **B**.[230] There are some similarities between the existence of extra dimensions of spacetime and the existence of the gauge potentials, although these are two different issues that relate to two different classes of theories. But extra spatial dimensions are related to gauge potentials and to the possible symmetries of the wave function.

[230] Most notably, the "New Localized EM Properties" was defended in the last decade by T. Maudlin and J. Mattingly (Healey 2007 55).

14.4. Symmetries and ground states: Kaluza versus Klein

I add now an important element related to the symmetry associated to the x^4 and let us return now to Klein's result. His theory does not apply to electron or to known any real particle. The suggestion is that the symmetry of electromagnetism **EM** and a rotational symmetry in the space of the wavefunction are in fact equivalent. Think of ψ as a complex number. Then we see that the transformation takes a wave function to another wave function and so forth:

$$\psi \to \psi' \to \psi'' ... \tag{146}$$

such that:

$$\psi'' = \psi' e^{i\beta}; \psi' = \psi e^{i\alpha}; \psi'' = \psi e^{i\delta} \text{such that } \delta = \alpha + \beta \tag{147}$$

From an algebraic point of view, this transformations of the phase factor form a group, the so called group of rotation U(1), i.e. the group of all unitary one-dimensional matrices (here, one-dimensional matrices are simply complex numbers **U** such that $\mathbf{U}\mathbf{U}^* = \mathbf{U}^*\mathbf{U} = \mathbf{1}$ and $\mathbf{U}^*$ denotes the Hermitian conjugate of **U**). It is a local group in the sense that all parameters in (156) can depend on the spacetime location. It is a Lie group which is Abelian (commutative), because all the transformations of phase commute (mainly because the sum within the exponent is commutative).

In this respect, the gauge group of **EM** is called U(1). Let us see now in what respect the U(1) group of **EM** can be recovered in Klein's theory. His insight was that the gauge group of electrodynamics U(1) coincided with the invariance on the compactified topology of x^4. Remember that the topology of x^4 is now a circle. The group associated with a circle is a group under multiplication, called the circle group $\mathbb{S}_1$.

In short, Klein recovered the symmetry group of electromagnetism by the **COMP** condition. To see this, we notice that the coordinates in 5-D can be altered only in one way: four translations on the first four directions:

$$x'_\mu = x_\mu + \lambda_\mu(x_\nu) \tag{148}$$

and a transformation given on the fifth axis.

$$x'_4 = x_4 + \lambda_4(x_\mu) \tag{149}$$

The tensor g_{mn} is transformed accordingly:[231]

$$\begin{cases} g'_{mn} = g_{mn} - \partial_\mu \lambda_\nu - \partial_\nu \lambda_\mu \\ g'_{\mu\nu} = g_{\mu\nu} - \partial_\mu \lambda_\nu - \partial_\nu \lambda_\mu \\ g'_{\mu 4} = g_{\mu 4} - \partial_\mu(\lambda_4) \\ g'_{44} = g_{44} \end{cases} \tag{150}$$

Indeed, x^4 is like an internal space or an internal degree of freedom associated to each point on $\mathbb{R}^4$. Later on, this type of internal space will be called "fiber" and the accompanying formalism is called the "fiber bundle formalism". The symmetries of electromagnetism emerge as geometrical consequences of the translation with a multiple of 2π in the fifth, "internal" direction. In Klein's theory, dynamical symmetries of the 4-D **EM** are taken as *consequences* of the geometrical symmetries of the compactified manifold; this is not only a unificatory aspect of Klein's result, but an explanatory one: internal symmetries are explained as external symmetries.

I mentioned that another problem that any unificatory theory needs to address is the compatibility between boundary conditions. Remember that Kaluza started from a Riemannian manifold in five dimensions. The space is a pseudo-Riemmanian space $\mathbb{V}^5$ asso-

[231] Here, as before, Greek indices run between 0 and 3 and Latin indices between 0 and 4.

ciated to a manifold (see Appendix for a definition of the Riemannian space). The ground state of such a space, the energy of a space in which there are no excited fields at all would be $\mathbb{M}^5$; but this is obviously not the case of our real world: we observe the ground state of a 4-D Minkowski space $\mathbb{M}^4$. If one writes the action in 5-D, (97) one expects the ground state to be $\mathbb{M}^5$, i.e. the Minkowski space. In general, unlike the ground state of quantum systems, the ground state in **GR** is problematic because of boundary conditions (Witten 1981b 414).

Setting this aside, one can still venture in analyzing the ground state of the two theories. In Klein there is a change of topology: we deal here with a manifold $\mathbb{V}^4 \otimes \mathbb{S}_1$ which has a ground state $\mathbb{M}^4 \otimes \mathbb{S}_1$. This is warranted especially because of the factorization of action (97). Several authors have showed that the ground state $\mathbb{M}^4 \otimes \mathbb{S}_1$ make more sense than $\mathbb{M}^5$ (Blagojevic 2002 295-299; Witten 1981b; Witten 1981a).[232] In fact, unlike Kaluza, Klein intuited that the "best" symmetry is $\mathbb{M}^4 \otimes S_1$. The symmetries of Klein's theory is the **4-D** Poincaré symmetry of $\mathbb{M}^4$ and a U(1) group of rotations of the circle $\mathbb{S}_1$. So the symmetry of Klein's theory is $P^4 \times U(1)$, while the symmetry of Kaluza's theory is P^5. The observed ground state is as measured is associated to P^4. This result has leverage in the explanatory power of Klein's theory because from these symmetries one can infer massless modes that correspond to gauge particles: the spin-two graviton and spin-one photon.

Despite Klein's serendipity, he tried to achieve too many things in a single strike: he managed to achieve some goals while he completely missed others. On the other hand, he has not been able to realize some goals for the simple reason that the time was not ripe

[232] What persists in all these recent analyses is the confusion between Kaluza and Klein.

yet. What is central for the present purposes is to take a look at the explanation present in both Kaluza and in Klein as a consequence of the unification.[233]

Unification and extra spatial dimensions in String Theory. This procedure has been manifest in the generalization of Kaluza-Klein to Yang-Mills field and later in String Theory: "our spacetime may have extra dimensions and spacetime symmetries in those dimensions are seen as internal (gauge) symmetries from the 4-D point of view. All symmetries could then be unified." (Ortìn Tomás 2004 291). This is the "Kaluza-Klein symmetry principle". Among other meanings, string theorists use unification as reduction of the *types* of symmetries.

What is the philosophical import of this result? Klein needs only the symmetries of spacetime. His theory is more parsimonious and simpler from the point of view of symmetries. We saw that a wave-function invariance demands geometrical transformations associated to the coordinates in 5-D. This type of unification reflects in letter and in spirit the creed of the "geometrization" program. The number of *types* of symmetry is then reduced, and not the sheer *number* of symmetries. This aspect of Klein's theory nicely echoes Kitcher's critique of Friedman's account of unification *qua* explanation: similarly, what matters here is the type of symmetries, not their number. For a bunch of reasons, in the programs inspired by geometrization, spacetime symmetries are preferred to internal symmetries.

---

[233] According to Pais, Klein met Pauli in 1927 to drink a bottle of wine "on the death of the fifth dimension". P. M. Dirac also advised Klein that his main trouble came from "trying to solve too many problems [to wit, the geometrization of electromagnetism as well as of the quantum theory] at a time," Pais claims also that Klein repudiated all he had written on this subject since 1927 and adds that "Modern string theorists, who believe that many dimensional theories will lead to the holy grail, may like to reflect on Klein's change of heart" See (Pais 2000 133). I suppose Pais refers here to a literal interpretation of a fifth dimension for which we have no empirical evidences. For me, this does not mean that extra dimensions of spacetime are "dead": either we'll see clear signatures of them later or we do literally not interpret them as dimensions of spacetime.

In addressing question [64] (p. 301), one might ask whether Klein's theory illustrates a reduction or a unification. In fact Klein reduced the internal symmetry of **EM** to the external symmetry of spacetime. We have here two theories: the **GR** and the **EM**. What seem to be internal symmetries of theories in 4-D are symmetries of space-time in extra dimensions. This can be taken as another case of serendipity on Klein's behalf. This was not intended by Klein, but it was discovered subsequently. The nature of the main assumption **CYL** in Klein is topological, while the results are formulated either in the language of **QM** or in that of the group theory. Is this a favorable way of wedding **GR** with gauge theories (here **EM**) and with **QM**?

I avoid the glitters of the symmetry reduction simply because the world is more complex than this. Obviously enough, Klein did not transform the symmetry of **GR** in an external symmetry of spacetime. The symmetry group of **GR** has always been a troublesome topic. By Noether's first theorem, to each and every symmetry of a theory there is a conservation law (this is the case with **EM** and electroweak theory). But gravitation has no conservation, so its symmetry is difficult to describe. Namely, the diffeomorphism group has nothing to do with the symmetries of spacetime manifold. Gravity can exist in various manifolds having different topologies. Klein was not able to reduce this internal symmetry to an external symmetry of spacetime. The fact that time evolution is a gauge motion produces the famous problem of time in canonical **GR**. But what Klein showed is that the symmetry of the other theory, **EM** can become a symmetry of spacetime.

In a nutshell, neither Klein nor other geometrical programs were able to reduce the internal symmetries of theories to external symmetries of spacetime manifold. The very distinction is actually problematic in the case of canonical **GR**. What can be admitted is a weaker form of type reduction of symmetries in Klein's unification: Kaluza-Klein theories can reduce the internal symmetries of fields *other than gravity* to external symmetries of spacetime manifold with 4D+ dimensions.

14.5. Klein's novel type of unification and Kitcher revisited

In addressing [49] and [50] (p. 261) I want to raise two issues here. I argue that Klein's theory is farther from a conjunction than Kaluza's. As I showed, a uniform description of **EM** and **GR** is possible (Section 8.4). I claim here that Klein's theory is more than an integration of two theories in the same formalism. **EM** and **GR** can be brought together in several ways, the less trivial being the formalism of action when both theories are expressed within one and the same mathematical structure. More generally, lots of theories can be expressed in common languages and expressed in similar terms: think how powerful is the language of **ODE** or **PDE** in this respect: but neither of these two procedures constitutes unification. In few cases like these there is any explanation present. Integrating two theories in the same formalism or language does not explain and does not predict anything. What Klein achived is more than this and it is a novel form of unification. There are two reasons to believe that Klein's theory is not a mere integration.

*The **EM-GR** coupling*. Here is a first important aspect of unification in Klein: there is no more modularization and pure sectors in Klein's metric (112). The form of the new metric contains the interaction term with A_μ within $g_{\mu\nu}$ and not only on the exterior stripes as in Kaluza. It has no more a pure gravitational "piece". The interaction term in A_μ represents the coupling between gravitation and electromagnetism, on which Kaluza remained silent. This emphasis on the interaction has a major impact in the generalization of Klein's theory to Yang-Mills fields. Electromagnetic and nuclear interactions are described by gauge fields, albeit by different symmetry groups.[234] If a symmetry group of a theory can be in principle transformed into the symmetry group of a spacetime structure with D>4 then generalizations of Yang-Mills fields beyond **EM** theory are possible. The standard Model of Elementary particles was totally separated from the theory of gravity. It can be said that in the context of the generalization of the Kaluza-Klein theory to Yang-Mills field the unificatory power was even more spectacular than in the original incarnation of the theory.

In respect of the general accounts of unification, Klein's argument acts like an argument pattern in Kitcher's sense (see Section 3.3). Klein illustrates the problem of Friedman's account in counting brute facts. But both my case studies are good illustrations—up to a point to Kitcher's account. I endorse the idea that Klein's theory acts like a unificatory pattern that was replicated and generalized to other cases than the **EM** interaction. There are two problems in applying Kitcher's argument to the Kaluza or Klein theories. First, Kitcher insists on the derivational parsimony as a condition of unification, perhaps as a normative condition to any unification. I struggled to find such a derivational parsimony in Kaluza or in Klein.[235] Klein's argument is a deduction but it unusual flexible in changing the premises with the conclusion. Klein in fact took **COMP** as an assumption and this became the basis of his explanations. I do not think that all explanations need to be

[234] The fact that all other interactions than the **EM** interactions are described by non-Abelian groups play a major role in the generalization of Klein's theory. This generalization was carried out be DeWitt, Kerner, Trautmann (Trautman 1970, 29; Kerner 1968, 143; DeWitt 1965).

[235] It would be interesting to explore the other option that runs against Kitcher's "deductive chauvinism" proposed by I. Halonen and J. Hintikka: unification as induction (Halonen and Hintikka 1999, 27-47; Weber 2002, 145-154) but this is beyond the scope of the present dissertation.

deductive and I do not see why we need deductive parsimony in order to achieve unification. The other major problem I see with Kitcher's general account of unification is that even in simpler case such as Klein there are several patterns that start from the same premises and reach the same conclusion. A very strong argument pattern is the application of the action principle. In my view it is not unificatory and not explanatory either. Kitcher would have difficulties in showing what the difference is between a formal or structural unification and more substantial cases of unification. I claim that Klein's theory help us understanding **EM** and **GR** as having a common origin in the geometry of spacetime. Last but not least, similar to other general accounts of unification, Kitcher's is not context dependent and this is a ubiquitous problem with any broad approach in philosophy of science. His account does not capture the central idea of a coupling element that unification creates between the two previous theories. I deem that such a coupling element is difficult to be captured by Kitcher's approach, but it is central to Mauldin ranking schema as well as to Morrison's approach.

More importantly than recovering Kitcher's account of unification, I claim that Klein's unification explanatory without being centered on deduction. I argue that some explanations are not consequences of the formalism so they are less deductive that Kitcher's approach would impose. Also, this case is even less related to the D-N model of explanation.

Why is the coupling element essential to Klein? The metric suggests that the g_{44} can be set to 1 and this helps to interpret transformation of the fifth axis in respect to the other four axis as a local U(1) gauge transformation which is the symmetry of **EM**. What is divided in parts or sectors in Klein is the action, which splits directly in gravitational and electromagnetic actions as well a kinetic action of the field ϕ. I take this strong coupling as playing here a double role: first it shows why Klein's theory is not a mere conjunction of **EM** and **GR**; second, it gives us a further reason to rank Klein higher in Maudlin's scheme.

The quantum factor in Klein's unification. Maxwell's unification of optics and electromagnetism may be read as a reduction of optics to electromagnetism but it is far from clear whether optics disappeared within the electromagnetic theory. One can ask whether optics survived Maxwell's unification. In many respects the answer is positive if one thinks of geometrical optics, more important for our everyday technology, than the electromagnetic theory of light. Also, results from optics changed the initial electromagnetic theory. Even if light is an electromagnetic wave, at the level of theory one can say why even Maxwell's example was not merely a reduction. In the same respect, electricity and magnetism were synthesized in electromagnetic theory, but none of them remained unscathed. Various authors suggested this corrective feature of unification, as Klein unification best illustrates.[236] Klein's unification involves elements for other theories that the unified theories. Because there is a novel element used in the mechanism of unification, unification has a corrective aspect. Klein's unification illustrates this corrective feature of unification better than Kaluza's. But this is a promissory note only. Klein's theory was able to predict indirectly the existence of the graviton. The idea that matter can be viewed as the consequence of geometry has a very interesting consequence far beyond something that Klein could anticipate: it is possible to correct Einstein‘s equivalence principle and to show that in fact in **5-D** there are two masses: an inertial and a gravitational mass which are equal when projected onto the **4-D** space (Wesson 2007, 254). So Kaluza-Klein can act as a dis-unifier for

[236] The way in which Newtonian mechanics corrected Kepler's law is discussed in (Forster 1988 88sqq.)

GR, or better as a correction to the fundamental principle of **GR**. Another correction as a direct consequence of the Kaluza‘s theory is that if one drops the weak field approximation, a non-gauge invariant term appears in **EM**. All these consequences look terribly exotic and many would not take them seriously.

Last but not least, there is a more philosophical correction that Klein (and Kaluza) suggested: they show that our concept of dimensionality is not as secure as it looks. Sometimes we can see that two theories which describe the same world by using different dimensions have the same or almost the same consequences. String Theory and various newer versions of Kaluza-Klein theories are such examples. Kaluza-Klein teaches us that given a specific structure of physical interactions and some restrictions on the scale of energy and/or length, or time intervals, one theory is the best description. For other restrictions and other interactions, another theory will better describe the world. As Craig Callender rightly pointed out, the question about the dimensionality of the world and how we explain it does not have definite answers: we may think we have them, but in fact we do not (Callender 2004).[237] If I interpret Kaluza's and Klein's unification correctly, the answers depend on what we want to account for, on how much science we put in our explanations, how unified our theories are or how unified we want them to be, and last but not least on how we look at the world, with what instruments, at what scale, etc.

A skeptic may still suspect that there is nothing added to Klein's theory than the conjuction of two theories plus the **COMP** assumption. If we put **EM GR** and maybe several parts of the QM, especially in its interpretation of matter waves, are we adding anything substantial more than a couple of dubious entities and ghost fields? The answer is that Klein's theory

[237] A more spectacular and more dramatic illustration of this situation is the so-called duality in String Theory, discovered by Maldacena in 1997, this duality cries for a philosophical analysis.

explains unexpectedly facts which do not fall into this narrow category of conjunction of theories and assumptions. Kaluza's theory on the other hand does look like a conjunction of theories wrapped in an extra dimensional formalism. The scope of this part of my dissertation is to insist on the differences between Kaluza and Klein. Klein explains and, under some caviats, predicts more than Kaluza does. But what is interesting is that causation is not the main part of this improvement. I claim that something else differentiates Klein from Kaluza: the physicality of the former compared to the formal approach of the latter. In Klein there is more than integration, uniform description, amalgamation and "bringing together" of several theories. Many theories can accommodate lots of phenomena in a simple and compact mathematical representation. Sometimes this can be expressed in terms of symmetry groups being included in larger groups, as wel'll see shortly. But none of these results are in themselves explanatory. Klein's theory is explanatory beyond it having some interesting consequences of the formalism. If there is a Kaluza-Klein type of unification, there is definitively a Kaluza-Klein type of explanation as I will argue in the following two chapters.[238]

[238] Many thanks to Bill Bechtel and Nancy Cartwright for making this point clear to me.

Chapter 15. Explanation, spacetime and dimensionality

Kaluza-Klein theories are spacetime theories based on the geometrization on more than four dimensions. In this chapter I show in what sense such a procedure is explanatory. Before entering the details of explanation in Kaluza and Klein, it is useful to enlarge the context of the discussion and ask some general issues related to the explanatory power of spacetime theories:

[83] *Geometrical explanations.*

[84] *Which element of a theory is explanatory? (A) The formalism itself? or (B) the spacetime structure?*

[85] *Is dimensionality open to interpretation?*

[86] *What and how can dimensions of spacetime explain?*

[87] *Some "bare models" of extra spatial dimensions and their explanatory power.*[239]

With respect to *[83]*, two spacetime theories are significant here: **SR** and **GR**. Since the time of the Leibniz-Clarke debate, a serious question has lingered around the spacetime theories. Can spacetime constitute the *explanans* of any kind of physical phenomena? At different periods of time and for different reasons, Leibniz, Mach and Einstein, at least around 1916, had held such a position. The same metaphysical problem haunted **GR:** the postulation of an unobservable entity, i.e. the spacetime, in order to explain observables, the dynamics of material particles and their relative motion. For Einstein in the 1920s, this constituted a "factitious causation" and it is epistemologically illegiti-

[239] By "bare model" I mean a minimal, but explanatory model of a spacetime to which extra dimensions are added.

mate because it infers observables from unobservables. For Einstein, writing about **GR** in 1916, **CM** involved the "factitious entity" of space and time in order to explain the accelerated and the rotational motion of bodies. Epistemologically, this direction of reasoning was flawed and Einstein claimed on several occasions that only point coincidences and only relative motion are real (Disalle 1995 320; Lorentz and others 1952 117; Earman and Norton 1987).

M. Friedman and H. Reichenbach both have suggested that the history of spacetime theories, from Newton to **GR**, can be seen as successive stages of *relativization* of motion (Friedman 1983; Reichenbach 1958 (1928)). For Aristotle and for many other pre-theoretical outlooks, some directions and orientations in space are privileged. Later on, Galileo removed the privileged directions and privileged states of motion. From an historical point of view, each stage added something and subtracted something from previous theories.[240] Alongside with relativization, explanations and pseudo-explanations came in.

Negative answers to [84] are easier to give. Still, some skeptics think that spacetime structures themselves do not explain. In the Leibniz-Clarke debate explanation was both sought and dismissed. Indeed, some features of motion, especially rotation and acceleration cannot be explained in purely relational terms, so they can be associated and explained by the absolute spacetime, argument suggested at least by Newton's example with the bucket.[241]

There are several types of argument against the explanatory power of **SR**, as suggested by Morrison: (A) a mathematical formalism is not able to explain in general; **SR** is

[240] As Disalle remarked, along different historical episodes, some components are metaphysical, some are epistemological (Disalle 1995 328-330).
[241] Mach would later on argue against the argument that acceleration is "real".

deeply dependent on the tensor-calculus, so it cannot explain. (B) **SR** is a principle-theory in the sense of Einstein's distinction between principle theories and constructive theories and as principle theories cannot explain, *qed*: **SR** does not explain.[242]

15.1. Formalism as *explanans* in spacetime theories

No formalism can explain. R. Feynman was suspicious about too powerful formalisms when he wrote that "whenever you see a sweeping statement that a tremendous amount can come from a very small number of assumptions, you *always* find that is false. There are usually a large number of implied assumptions that are far from obvious if you think about them sufficiently carefully" (Feynman, Leighton, and Sands 1989 II-26-1 my emphasis). Morrison suspected something similar in the case of **SR**. We need to add several hidden assumptions in order to get explanations from this theory: many are innocuous, some are controversial, some are maybe blatantly false. In acknowledging the unifying power of the tensor calculus used in **SR**, Morrison pointed out that this type of unification does not come with an explanation and it is based only on mathematical properties. In referring to the covariant **EM**, she argues for [5] and [6] (p. 111-111) that the Lorentz transformation helps us to calculate everything, but basically explains nothing. Can we infer that **SR** in tot is not explanatory? Definitively not. For Morrison, this is similar to the Lagrangian formalism in **CM** which is mainly a computational tool that lacks any explanatory power. But the postulates of **SR** claim that they explain.

> However, the Lorentz transformations are not explanatory in and of themselves; they simply provide the tools for relating different reference frames given the two postulates, the definition of simultaneity and the relativity of

[242] This argument is often used to dismiss the block view of the universe or to endorse presentism, see for example (Craig 2000, 287). I do not follow further this line of thought.

> length and time. The transformation equations are in some sense the embodiment of the two postulates [of **SR**], but are not explanatory of the ways in which systems are constituted, that is to say, they don't provide a reduction of space and time, or of electricity and magnetism; instead they show how to integrate them so that physical phenomena and systems can be treated in a unified fashion. In the Minkowski case, geometry and physics are not integrated in the way they are in Einstein's [**GR**]. In the [**SR**] the geometry of space-time determines a possibility structure, rather than an explanation of how a particular system travels through space-time; that is, there is no geometric explanation of phenomena, in the way that curved space-time explains gravitational force. Consequently, the unifying power of the tensor calculus defined on that space-time does not provide a physical explanation of the integration of electric and magnetic fields that extends beyond the relativity described by Einstein in 1905. […] [**SR**] played a unifying role for physics proper by specifying formal constraints for its laws. That specification gave rise to a more localized unity of electricity and magnetism as well as space and time (Morrison 2000 190-191).

Nevertheless, for Morrison **SR** played a unifying role by specifying formal constraints. **SR** is for Morrison a synthetic unity and not a reductive one: Maxwell reduced optics to electromagnetism one can't help but feel that despite the similarities in synthetic character, it differs from the unity present in the electroweak case where the parameter of the unification is free. **SR** exhibits a greater overall coherence and hence a greater unity, but no explanatory power.

Here are some weak points in Morrison's argument. First, let us admit for the sake of the argument that covariant **EM** does not provide explanations of any kind. The simplest question to ask is: what are the alternatives to this theory and which theory is more explanatory? Do other formulations of **EM** explain better? Does Maxwell's **EM** fare better than the **EM** tensor form? Morrison does not clarify this point. I think asking whether tensor calculus explains or not is not meaningful, but without exposing the alternatives, Morrison does not have enough logical space for the argument. The question is whether other theories that dispense of tensor calculus explain better. I showed that the descriptions we can build without tensors are mired with coordinates and this is a serious problem. Einstein acknowledged in his "Autobiographical Notes" the invaluable contribution of Minkowski's legacy. Before Minkowski, each law of physics needed to be tested for invariance by carrying out a Lorentz-transformation. The 4-D formalism expressed essentially through tensor-calculus simplified the Lorentz transformation and geometrized the procedure of testing for invariance by re-defining the Lorentz transformation as rotations in 4-D space (Brown 2005 131). Tensors are difficult to understand and sometimes unnecessarily complicated, but it seems they are a necessary evil, possibly the least evil of all evils. We want to avoid representations heavily dependent on coordinates. They are not part of the physical world: they are representations and they come with a specific theory or a specific worldview if you want. Should we trust those explanations more than explanations based on tensor? It is not my aim to discuss here what a more suitable explanation of **EM** based on something else than tensors, although such alternatives are available. I suspect that Morrison's analysis, besides its plausible criticism of "mathematical explanations", lacks a positive and constructive outcome. Her problem here is not the tensor formalism, but the lack of the machinery, i.e. the causal story. Her criticism is misplaced in the sense that what would explain is not the tensor calculus itself, but the spacetime structure. As I stated in Part I this causal reflex, our own obsession to find the causal structure can be misguiding in fundamental science: as far as I know Morrison never addressed this worry. She could discuss whether a tensor calculus can capture the *causal* story of the world *better* or *worse* than, let us say the three-vector calculus. That could be an interesting discussion. But reducing explanation to causation and SR to the tensor calculus is misleading.

Morrison mentions *in passim* **GR**, but nowhere in the book has she discussed **GR** as unificatory theory. She does not show what the unificatory power of **GR** was, whether it was related to the formalism of the tensor calculus or not—a discussion of **GR** is almost completely absent from Morrison's analysis. Is it that tensor calculus unifies in **GR** but it does not unify in **SR**? **GR** is the standard bearer of theories using tensor calculus. For all these reasons, I think that her criticism of **SR** is incomplete at best.

The broader context to which my analysis belongs is question #56. If Morrison is right, mathematical formalisms and theories deeply dependent on formalisms do not explain, although they have consequences and results. How do differentiate results or consequences of a specific mathematical hypothesis from genuine explanations? Is mathematics explanatory? In general, I think it is not and there are trivial explanations in mathematics as well as in physics. I give an answer to these questions and to those related to #56 in the case of Kaluza and Klein. Morrison is wrong in claiming that formalism cannot explain. Although I take Kaluza's theory as a formalism which has some relevant mathematical consequences, I admit that it does not explain or at least it has a very limited explanatory power. On the other hand, Klein's theory explains physical phenomena. I envisage Klein's theory more as a mixture between consequences of the formalism itself and explanations which come from the formalisms and the *physical* assumptions the theory makes.

Before discussing the details of Klein's theory it is useful to expose some arguments that pertain to [56] and [57].

15.2. Is Minkowski spacetime properly an *explanans?*

Principle theories do not explain. Instead of blaming the lack of unificatory power on the formalism itself, it is more appropriate to discuss the explanatory power of space-time structure as related to the source of the theory. In the last years, Einstein's distinction between principle theories and constructive theories has drawn the attention of philosophers, again. I do not want to enter into the details of Einstein's distinction. Einstein explained it very briefly in an article wrote for *The Times* (London) in 1919: a constructive theory starts from a relatively simple formal scheme and build up a picture of more complex phenomena. Constructive theories are built from formal hypotheses. The more complex data we gather are the results of constructive theories. According to Einstein, constructive theories are most common. They attempt "to build a picture of complex phenomena out of some relatively simple proposition" or out of hypothetical constituents. Einstein's example of constructive theories was the kinetic theory of gases**.**

Principle theories start from the phenomena and from evidences and then build the formal structure of the theory. Thermodynamics is definitively such a principle theory and for Einstein, the 1905 formulation of **SR** was a principle theory, too (Brown 2005 71; Balashov and Janssen 2003, 327-346). A principle theory depends upon the observations of the phenomenon—its empirical content one can say. By empirically generalizing from phenomena, one builds a theory that applies to "every case which presents itself". Thermodynamics takes the observation that we never find perpetual motion in nature as its principle. This becomes then part of the first and the second law of thermodynamics. In fact classical geometry with its Euclidean metric and Euclidean space is fundamentally phenomenological. It takes some of the observable properties of the physical space such as angles and distances among bodies and promotes them to axioms or postulates.

Is SR explanatory deficient? Some authors dissent to the claim that all principle theories are explanatory deficient (Craig 2000 109) and that a constructive formulation of a principle theory is not always possible. For example, D. Howard writes that "ultimate understanding requires a constructive theory" and suggested that principle theories serve to constrain the search for constructive theories (Howard). Principle theories are only constraints acting on constructive theories. H. Brown and C. Timpson suggest that we can dispense from the beginning with principle theories and develop the right constructive theory from the get-go. Principle theories are the last resort used when constructive theories are either unavailable, too difficult to build, or too awkward too understand.

If this is the case, one question is whether there is a constructive version/formulation of a principle theory. This question was asked for **SR** by several authors (Brown 2005 132). For M. Friedman for example, Minkowski geometry added to the **SR** is needed to the constructive version of **SR**. The Minkowski structure of spacetime is ontologically autonomous and needs to be added to the **SR** theory in order to explain the behavior of complex material bodies, independently of their material constituency (Friedman 1983 Ch. 4). The opposite view has been advocated by H. Brown and O. Pooley: the explanation of, let us say, the length contraction, is ultimately sought in terms of the dynamics of the microstructure of this specific contracting rod: the material constituency of clocks for example counts when we want to explain why time dilates. This is called sometimes the Lorentzian or Neo-Lorentzian pedagogy for example by J. Bell and Brown and Pooley (Brown and Pooley 2001 257). The *structure* of the bodies used to measure space or time, i.e. rods and clocks and atomic or nuclear device we usually employ to this end matters. It must not be overlooked by the analysis of spacetime metric. According to Balashov and Janssen, to explain the geometrical structure of spacetime in terms of invariance of forces means to put the "cart before the horses" (Balashov and Janssen 2003 340-341). The Neo-Lorentzian confuses the causes with the effect and takes the *explanandum* for the *explanans* and vice versa.

It seems that for the (Neo-)Lorentzians clock and rods does not fully capture the significance of the metric field specific to **SR**. The point made by Weyl for example in his answer to Einstein's objection was that one could not know how a clock would behave under high accelerations, or intense **EM** field or any other extreme conditions as long as one did not have a "full" dynamical model of the clock. But this bring us back to the duality between matter and fields and points towards the idea that explanatory power of the **SR** theory the fact that it can practically explain the length contraction and time dilation is not due exclusively to the metric and its form, or to the presence of fields of any type, but to the measuring rods and clocks, to their internal structure and of course ultimately to their quantum mechanical properties. Einstein in 1949, in *Autobiographical Notes*, deplored this situation and wrote that it was a "sin" to take measuring devices like rods and clocks as irreducible (Schilpp and Einstein 1949 59, 61). In other words, geometry is not fundamental or not enough to explain even simpler facts as the length contraction and time dilation. More extremely put, spacetime theories are essentially matter theories.

A critical note on the distinction principle/constitutive. Notwithstanding Einstein's own quick treatment of this distinction that reflects his own specific approach to the philosophy of science, it is debatable whether **SR** is a pure principle theory or not. In the same vein, it is not clear whether **QM** is a constitutive or a principle theory.[243] I suspect the distinction is questionable on several grounds. As everybody would admit, there are no pure constructive or pure principle theories. Moreover, the evolution could move a science from being a principle theory to a constructive theory. Third, I think the most serious problem is the theory-ladeness of observation. All instruments of measurement of time or space are based on some interaction or other. This suggests that the distinction in itself is highly operationalized. Or even if the distinction can hold water at least in some cases, it definitively cannot mark the distinction between explanatory and non-explanatory theories.[244]

Explanation and Minkowski spacetime structure. It seems the Minkowski spacetime, with its simpler structure, does explanatory job. It explains simple things such as the time dilation and length contraction, it explains why objects with real mass (not imaginary mass) cannot be accelerated to speed above the speed of light, explains what several observers in inertial. The advocates of the causal explanations in science would bark to the idea that there is explanation where causation is absent. Let me ask a blunt question: is causation present in **SR**? I suspect that causation is not in fact fundamental in **SR**, but it is a different way of describing time, Minkowski spacetime and possibly the *ordering* relations among objects. Although ordering and causation are two different things, one can see how the gap between **SR** and causation can be bridged together. The idea that we have on one side tensors, **SR**, formalism, mathematical trickeries used to unify and so on and on the

[243] J. Bub advocates **QM** as a principle theory, while H. Brown, C. Timpson, A. Hagar, advocate the constructive part of **QM**. Seemingly, the debate has intensified after Bub's article (Bub 2000). Several authors claim that the information theoretic formulation of **QM** is its principle formulation (Clifton, Bub, and Halvorson 2003).

[244] There are several authors who challenged this distinction.

other side, causation, explanation, the machinery, understanding etc. seems to me too simplistic a philosophical analysis.

The previous argument of Brown and the opposite view, according to which the geometrical structure is the *explanans* ((Balashov and Janssen 2003)) suggest that both the geometrical structure and invariant forces acting on matter can play the role of *explanandum* and of *explanans* likewise. Then what causes what? The answer here is that causal and non-causal or structural explanations can coexist and provide different path of explanations to the same phenomena. Admitting a plurality of types of explanations opens the possibility to have different explanations pointing in different directions.

One word about my own case study: Kaluza and Klein tried to relate the form of the **EM** interaction to the properties of the spacetime structure without suggesting a causal connection among them, although such a connection is possible. Klein moreover struggled with the causal story. First, both adopted **GR** which is not clearly causal in its full generality. Second, Klein's **COMP** condition forced him to think in terms of incredibly small distances, roughly the size of the Planck scale.

15.3. Non-causal explanations and spacetime

Somebody who admits that there are other types of explanations than causal explanations can ask which of them is the most relevant (Lipton 2004 31). First, the causalist has to be convinced that there are non-causal explanations. For example there are mathematical explanations which are simply not causal: functional dependencies, mathematical truths, explanations based on symmetries, etc. For example, explanations based on symmetries are non-causal.[245]

One of the most vocal advocates of spacetime properties as *explanans* is G. Nerlich. He tries to answer questions such as: "If X is a property of spacetime, what does X explains?" In his analysis, X is constant or variable curvature, handedness of spacetime etc. (Nerlich 1994 Ch. 7) and a question less relevant for the present case "Is X reducible to some events involving matter?". Nerlich shows that curvature is explanatory without being causally related to events involving matter. Roughly speaking, I ask a similar question where X is the dimensionality of spacetime.

The most charitable solution is to endorse explanatory pluralism and accept that the causal account of explanation is not complete. It is also perfectly ok to compare causal and non-causal explanations when they compete and look for the best explanation, causal or otherwise. The strategy of expanding the causal model over its own limit is not advisable. There were authors who tried to follow this route. Others tried to include "determination" in the causal account of explanation (Ruben). More recently, M. Strevens tried to extend the causal account of explanation such that it can provide all what unificationist can give (Strevens 2004, 154-176).

In comparing non-causal and causal accounts, Lipton concludes that:

> For the time being at least, I believe that the causal view is still our best, because of the backward state of alternate views of explanation, and the overwhelming preponderance of causal explanations among all explana-

[245] Lipton's example is simple: throw a bunch of sticks in the air and let them fall. Take a snapshot any moment: you will see more sticks closer to the horizontal position than sticks which are almost vertical. The explanation is simple: in D=3 there are infinitely many horizontal positions, but only two vertical positions. This is due to a symmetry imbued in our usage of the words vertical and horizontal: there are two horizontal directions, but only one vertical direction.

> tions. Nor does it seem ad hoc to limit our attention to causal explanations. The causal view does not simply pick out a feature that certain explanations happen to have: causal explanations are explanatory *because* they are causal (Lipton 2004 32).

I do want to push further the issue of causal versus non-causal explanations. I do not want to ask where we should look for non-causal explanations, but, better, where we can lay down the field for a confrontation between causal and non-causal explanations. If there is room for non-causal explanations, are there reasons to prefer one over the other?

Reification of spacetime structures: the gauge case. There are also current debates on whether geometrical structures similar to those used by the Kaluza-Klein theories play any relevant explanatory role. Gauge structures and their geometrical or topological properties are at the center of such a debate (for a general discussion, see (Leeds 1999, 606-627; Healey 2007, 297; Gordon Belot 1998, 531-555; Healey 2001, 432-455; Batterman 2003, 527-557). Another question that arises in the context of gauge structures is their reification. Sometimes a genuine explanation of certain phenomena requires an appeal to geometric or topological features which are related to abstract spaces. It seems that although gauge structures do not act causally they are explanatory nevertheless.

Last but not least is a question related to the completeness of the geometrical explanations. What is needed besides topology or geometry in order to ascertain explanation in gauge theories? In the case of the Aharonov-Bohm effect for example, Healey shows that basic physical principles have to be added to geometrical or topological considerations. In some cases, these principles need themselves interpretations. For Healey, entering into interpretative debates around **EM** is necessary in order to understand the Aharonov-Bohm effect (Healey 2007 41-42).

15.4. Interpretations of the theoretical structure in Kaluza and Klein

How are the debates about the interpretation of gauge invariants related to Kaluza-Klein theory? I claim that similar philosophical questions can be asked here and that the answer to [85] is affirmative. There are reasons to entertain such a parallel discussion. First, gauge invariance and Kaluza-Klein are similar. Second, fiber-bundle as a formalism of gauge theories and its interpretation is related to Kaluza-Klein. Third, there is a sense in which the fifth dimension added in Kaluza-Klein is "like" an internal degree of freedom. In other sense, it is less than a physical dimension but more than a trick of our representation through coordinate system. Given the extreme difficulties to measure or observe it, the fifth dimension has had this ambiguous interpretation even since Kaluza and Klein. Interpretative considerations are in order here as they are in the case of the gauge theories. String Theory also is open to similar interpretations.

Interpretation of the theoretical parameters is present in both incarnations of this theory. While philosophers like theories open to interpretation as prolific fields for philosophical investigation, scientists prefer to give clear physical meaning for any theoretical term. As I showed, Kaluza and Klein theories have uninterpreted theoretical terms. A formal result such as the Campbell-Magaard theorem does not come with any interpretation of what happens in the embedding space. Kaluza's idea was to add a spatial dimension as directly related to the **EM** interaction. Denying the hypothesis that spacetime has exactly four dimensions is a bold claim, although it is clear that all the arguments pertaining to show that D=3 are also restrictive and incomplete (Callender 2005, 113-136). Kaluza and Klein denied that extra spatial dimensions are similar to the three spatial dimensions we know—an assumption on which arguments for D=3 are based upon. This is not the case in Kaluza and in Klein. What differs is the interpretation of the fifth component of the impulse-energy tensor or, in Klein, the topology of the fifth dimension. Both Kaluza and Klein were vague in their interpretations, which constituted the reason to dismiss the theory; they were in fact ambiguous in providing physical interpretations to the structures and operations on the fifth direction. Also, the theory had difficulties in representing matter fields in **5-D**. For empirical reasons, it was obvious that extra dimensions need to have a special feature when compared to the other three spatial dimensions. The idea of representing forces, or later on, fields, in extra-dimensional spacetime in order to unify them is one of the common features of Kaluza-Klein and String Theory. Similar to String Theory, extra spatial dimensions are in fact special. This makes Kaluza's and Klein's theories something more than "mere conjunctions".

Mathematically, extra dimensions are not bad in themselves and they were used extensively in classical and statistical mechanics long before Kaluza. All discussions jump immediately to the question of the reality of the extra dimension added to the formalism: is it real or it is a feature of the representation? As I mentioned, reality check has been performed in different contexts such as gauge theories (are gauge potentials real?) or **QM** (is the wave real or just a feature of our representation of the world?). Even in **CM** one can ask how real the configuration or phase spaces are or any representational space for that matter. This was too incipient a step to take in the case of Kaluza. In avoiding questions about scientific realism, I entertained tacitly the middle ground: the extra dimension is not spatial in the same sense as the other three spatial dimensions are. If it is real, it is not real in the same way x^1-x^3 or x^0 are. In adopting this middle position, I have some leverage in my argument: the dimensional extension from **3-D**+time to **4-D** spacetime by Minkowski, Lorentz, Einstein, etc. When Minkowski added time as a fourth dimension to the existing three spatial dimensions, some features of the new dimension were clearly not spatial. Time is not more real or less real than the spatial dimensions; it is real in a different way… One can reply that the transition **3D+1** to **4-D** has clear and testable empirical consequences. In fact, there is no empirical evidence against or for extra spatial dimensions. This is indeed the difference that makes Kaluza and Klein a theory open to interpretation.

An interpretation of the x^4 needed to be added to Kaluza-Klein theory. Another question to be asked here is whether we need a high dimensional *manifold* associated with the high dimensional tensors. Is there a way to talk about high dimensional tensors in lower dimensional manifolds? The projective geometry program carried on by Veblen, Hoffmann, Bergmann, Einstein *i.a.* assumed that we do not need a real **5-D** manifold, although we can populate the 4-D manifold with five-dimensional tensors, vectors or even coordinates. But in Kaluza's and in Klein's theories, there is a **5-D** manifold on top of the multi-dimensional tensors. We'll see how this embedding in a higher dimensional structure has unificatory power.

Realism about extra dimension. As always, philosophers would like to push unification further than the boundary of epistemology. Once unification is achieved, one can ask the same question as Friedman asked about spacetime structures: do we interpret *literally* the unificatory structure or on the contrary, we keep it at the level of representation, i.e. we map it partially onto reality? Unfortunately, for the present case study, there is no definite answer to such an extension of unification toward realism. On the existence of extra spatial dimensions, the jury is still out—the same holds for the discreteness of space and time, finitude of space and time, etc. and

Then was it the fifth dimension of spacetime? Is it real or not? The present dissertation is not proposing a definite answer to this question but only a guide in answering it. Most probably, the fifth dimension, if it exists at all does not look as simple as Kaluza and Klein depicted it. But there is a more positive answer to this question: the fifth dimension is the best representation that unifies **GR** and **EM** in classical context and given such and such assumptions and approximations. It also provides intended and unintended explanations and some predictions. It also acts as a template for more elaborate and more accurate models of reality. Moreover, it illustrates unification.

Another unanswered question relates to the reality of the field ϕ. There is definitively a difference between the field ϕ and the other components of the metric. One can press for the same, essential question: what is real in Kaluza's or Klein's **5-D** world? We can ask a simple question similar to the one in gauge theories or in quantum mechanics: what is physical? I see here three options, similar to the discussion about the reality of gauge potentials:

- The g_{mn}, i.e. the **5-D** metric field.
- The $g_{\mu\nu}$ and A_μ, i.e. the 4-D metric field and the **EM** 4-vector and possibly ϕ, too;
- A possible integral structure living in the **5-D** spacetime.[246]

If we adopt the first option we need to think of ϕ as real or being as real as $g_{\mu\nu}$ and A_μ. Otherwise we have to face a strange situation in which some components of the metric tensor are less real than other components. This definitively creates an ontological

[246] This can be the closest interpretation of Kaluza to what is in today's parlance String Theory where string or better branes live in the compactified dimensions.

asymmetry hard to adopt. If we adopt the second option, than we are closer to a mere conjunction.

Does the x^4 parameter have the same ambiguous regime as the gauge potentials? The battle fought on the reality gauge is relevant in the context of extra spatial dimensions. Needless to say that no matter how powerful the analogy is, one cannot simply infer the existence of extra spatial dimensions from the existence of gauge potentials or the other way around.

15.5. Are false theories explanatory?

Although Klein's theory unifies and explains, almost surely it is not true in a literal sense. Doesn't a purported explanation have to be true to be explanatory? This is the case, for instance, according to Hempel's **D-N** model of explanation. If Kaluza's or Klein's theories are very likely to be false, I have a problem here, haven't I? I have a couple of answers. First, I'm not assuming the **D-N** model of explanation, so even a false theory can explain. If the reader likes, she can understand this case study as illustrating the distinctive benefits to understanding provided by Kaluza-Klein, if it were true. Other more notorious unificatory theories were false. Lots of successful and explanatory theories have been proven to be false. There is philosophical work looking at the role of unification in Newtonian physics despite the fact that Newtonian physics is wrong. More importantly, and interestingly, although Kaluza-Klein is false, its pattern of explanation lives on. The theory itself is malleable enough to adopt forces other than electromagnetic interaction.

There are other, however, severe limitations of the theory, though. It cannot accommodate chirality, a fundamental feature of particle fields, and it has difficulties in representing fermions. Kaluza-Klein theory is definitively not the fundamental theory for at least one reason: it unifies only two classical theories which are not the fundamental theories. And in fact there are many current theories that are, broadly speaking, Kaluza-Klein-type theories. If one of these is correct, then Kaluza-Klein might be seen as in some sense approximately true. This whole dissertation is in fact about a strictly false theory.

Philosophers analyze exotic objects and hypotheses even if they are strictly speaking false. Looking for Closed Timelike Curves is in its own an industry in **GR** and in the philosophy of physics (Earman, Smeenk, and Wüthrich 2009)

. Even if they do not exist, they tell us a lot about how our world is or about what the laws of the universe are.[247] Or weaker, they tell us important things about the theories by which we represent the world if not about the world itself. The same can be said about exotic topologies, rotating universe, **GR** in D=2, etc. All of these are interesting in themselves. In 1949 K. Gödel studied a rotating universe in this context—even if there was no empirical reason to believe that the universe was rotating or that it wasn't. This led him to (re)discover Closed Timelike Curves. A different argument for why these exotic entities may interest us is based on probability: we can estimate the probability that a close timelike curve could exist in such and such universe, even if there are none. At the end of the day, we cannot rule out Closed Timelike Curves by *fiat.* The study of possibly non-existent object is a reputable scientific task in itself.

[247] **QM** is non-relativistic, **GR** doesn't talk about matter, the standard model ignores gravitation etc. All our theories are in fact lying in some degree. Thanks to Christian Wüthrich from making this point clear to me.

I think that the hypothesis that our spacetime may have extra spatial dimensions falls under the same category of “exotic” hypotheses. The purpose of the present analysis is to argue that such a hypothesis has explanatory and unificatory power, even if it is literally false. There are other highly unified theories which are false but increased our knowledge about the world: Maxwell is such an example and Kaluza-Klein qualifies here as such. On the other hand there are true theories which say nothing about the world and do not improve our knowledge of it and that are highly irrelevant. What is important here is to draw the lessons from analyzing false, but relevant theories that say something about the world or about how the world might be.

Even if we live in a universe with exactly three spatial dimensions (D=3), the analysis of D>4 can reveal some counterfactual aspects of the actual laws which are impossible to reveal otherwise (the same can be said about an analysis in D=2). On the other hand, we still do not know whether our universe has precisely three spatial dimensions (D=3) everywhere and at anytime. The skeptic could reply that a non-testable theory would not be very significant either. We need to adopt an epistemic humbleness in this respect and admit that we know too little about the real dimensionality of space and time. Some general results of my analysis can be useful to a class of theories in theoretical physics, even if the theory under scrutiny is literally false. We can also change the perspective here and talk about Kaluza-Klein as an incomplete theory of which we do not know whether it is false or true, but of which we know it is incomplete.[248] If String Theory will be disconfirmed in the future, Kaluza-Klein theories will become less relevant to the philosophers of

[248] Thanks to Ken Intriligator for clarifying this point.

science. But if a version of String Theory will be confirmed in the future, the present dissertation will play a role in the philosophical approaches.

15.6. An excurse in "Hyperspace"

It is important to explore the logical space of explanation of the assumption that the physical space has extra dimension(s). False scientific theories about extra dimensions of spacetime have a long history and frequently they intermeshed with science fiction. The answer to [86] is: "almost anything". The history of extra spatial dimensions is interesting in itself as it is linked to explanation and pseudo-explanations. When extra dimensions of spacetime are not related to physical interactions or to a scientific theory about matter or fields, any type of unexplained and unknown fact in the world can be related to extra spacetime dimensions. Needless to say, almost all such endeavors end up as pseudo-explanations. We have empirical evidence of the dimensionality of spacetime, more precisely of the fact that there are *at least* three spatial dimensions because that's how the world appears to us. The philosophical question asked since Aristotle and Euclid is:[249]

> ***[88]*** *What is the dimensionality of physical space? Is it three (D=3) or greater than three?*

Although since recently physicists have avoided playing with the dimensionality of spacetime, arguments for and against the necessity of D=3 were produced since Antiquity. Philosophers and mathematicians approached this problem and some results were already known in the 19th century:

[249] By the way of a convention, hereby I designate the dimensionality of *space* by the letter D; when time is involved I use a hyphen to designate various types of *spacetime* manifolds: 4-D or 5-D.

- D=3 (Aristotle, Pappus, Ptolemy):[250] space has D=3 dimensions because we cannot conceive extra dimensions of spacetime;
- D=3 (Kant): space has D=3 dimensions as a consequence of the law of universal attraction (Newton's gravitation law).[251]
- D=3 because some orbits (planetary or atomic) are stable only if D=3 (Ehrenfest, Barrow, Büchel; for a criticism of these arguments see (Callender 2005))

---

[250] A complete treatment of Aristotle's philosophy of space and time would take too much. I reproduce here his argument against hyper-dimensional bodies: "Now a continuum is that which is divisible into parts always capable of subdivision, and a body is that which is every way divisible. A magnitude if divisible one way is a line, if two ways a surface, and if three a body. Beyond these there is no other magnitude, because the three dimensions are all that there are, and that which is divisible in three directions is divisible in all. For, as the Pythagoreans say, the world and all that is in it is determined by the number three, since beginning and middle and end give the number of an 'all', and the number they give is the triad. […] body alone among magnitudes can be complete. For it alone is determined by the three dimensions, that is, is an 'all'. But if it is divisible in three dimensions it is every way divisible, while the other magnitudes are divisible in one dimension or in two alone: for the divisibility and continuity of magnitudes depend upon the number of the dimensions, one sort being continuous in one direction, another in two, another in all. All magnitudes, then, which are divisible are also continuous. […]
One thing, however, is clear. We cannot pass beyond body to a further kind, as we passed from length to surface, and from surface to body. For if we could, it would cease to be true that body is complete magnitude. We could pass beyond it only in virtue of a defect in it; and that which is complete cannot be defective, since it has being in every respect." *De Caelo*, 1,2. 268b (translation by J.L. Stocks).

[251251] "It is probable that the three-dimensionality of space derives from the law according to which the forces of substances act on each other. Because everything found among the properties of a thing must be derivable from what contains within itself the complete ground of the thing itself, the properties of extension, and hence also its three-dimensionality, must also be based on the properties of the force substances possess in respect of the things with which they are connected. The force by which any substance acts in union with other substances cannot be conceived without a certain law that manifests itself in its mode of action. Since the kind of law by which substances act on each other must also determine the kind of union and composition of many substances, the law according to which an entire collection of substances (i.e., a space) is measured, or the dimension of extension, will derive from the laws according to which the substances seek to unite by virtue of their essential forces. The three-dimensional character seems to derive from the fact that substances in the existing world act on each other in such a way that the strength of the action is inversely proportionate to the square of the distances. Accordingly, I am of the opinion that substances in the existing world, of which we are a part, have essential forces of such a kind that they propagate their effects in union with each other according to the inverse-square relation of the distances; secondly, that the whole to which this gives rise has, by virtue of this law, the property of being three-dimensional; thirdly, that this law is arbitrary, and that God could have chosen another, e.g., the inverse-cube, relation; fourthly, and finally, that an extension with different properties and dimensions would also have resulted from a different law." (*Gedanken von der wahren Schatzung…* (1746), translated by Eric Watkins. Thanks to Eric for making this translation available to me before it was published.

- D>3, the Hyper-space or the hidden dimensions hypothesis: Space has more than three spatial dimensions in reality, although it appears "as if" D=3. Extra spatial dimensions are hidden because of several reasons.
- D=3 is neither true nor false, it is conventional (Poincaré). In other words, the dimensionality does not matter that much and can be postulated at our convenience.

What is also at stake is the modal character of the claim D=3. Is it a necessary or a contingent truth? Could it be the case that the world could have more than three spatial dimensions although in reality it has only three? Is D=3 accidental or by necessity?[252]

D>3 was always associated with occultism and mysticism. Einstein wrote in 1961 that people without a mathematical education are seized by a mysterious shuddering when they hear of space with more than four dimensions (Einstein 1961, 164). It is no wonder that this attitude is reflected in today's criticism against String Theory. For many physicists, a book about D>4 is like a book about unicorns or on Santa Claus. Here we need to step ahead of the problem that we do not have evidences that we live in a universe with D>3. In foundational studies, philosophers may need to deal with infinity, possible worlds, multiverses, time travel, discrete time or space, wormholes, twistors, loops, strings, branes, which are after all (maybe) non-real entities, but modes of our representations of the world. Extra spatial dimensions or the hyperspace may well fall under this category. It would not be for the first time that non-existing objects have major importance in the development of science: even the idea of time as the fourth dimension was initially a pure speculation and

[252] This modal character of answers to [88] is indeed a very challenging philosophical question that stirred philosophical interest. If my interpretation to Kaluza and Klein is correct, the dimensionality of the world is contingent upon the interactions we have in the world. I will come back to this issue later.

now we are accustomed to use it in fundamental physics. Entries in the above list may well be as real as elementary particles or as unreal as the phlogiston and this opens a host of arguments for anti-realism and this is not my intention.

The whole literature on hyperspace is fueled by imagination, by some philosophical thought experiments or by the sheer science fiction. The analogy used here is simple. If we can fold a D=2 surface (a piece of paper) in our D=3 space, then our 3-D space ("land" or universe or whatever) can be folded in D=4 etc. Early sci-fi literature on hyperspace depicted hyperspace as an apple. Our ordinary space is the surface of the apple and we can travel from one point to another by several paths *on* the apple. But there will be always shortcuts *through* the apple, for example from one pole to the other. The suggestive term "wormhole" was coined by J. Wheeler and C. Misner in the 1950s when they wanted to show that electricity can be understood as force lines trapped in such wormholes (Nahin 1993 81; Misner and Wheeler 1957, 525-603). Interestingly enough, **GR** itself does not prohibit the existence of such shortcuts through the apple, fact known to the **GR** community as early as 1916 (L. Flamm studied this possibility). It is true that the so-called Einstein-Rosen bridges in spacetime can be ruled out on other considerations which are now clearly related to **QM**. Bluntly put, these bridges would be incredibly instable, if they are possible at all. But they are excellent heuristic tools to understand the theory of **GR** itself and its stability to these heretic spacetime structures. Studying wormholes, extra dimensionality or gravity in D=2 (which is blatantly not the case of our real spacetime) can have at least an impact our understanding of the theory in D=3. In fact, science deals with counterfactual situations, not only with reality.

The suggestion is that extra space dimensions are “special” in some respect. What is special about the extra dimensions of space? This is the core story of the literature on hyperspace. Before stepping into the idea of extra spatial dimensions related to physical interaction, let us look in the early literature on hyperspace which is not related to physical interactions.

Transcendence and moral dimension. I mentioned several times that extra dimensions of spacetime can be used to explain almost everything. There is a clear trend to associate extra dimensions of space and time to spiritual or immaterial objects. Henry More (1614-1687), the Cambridge Platonist, associated a fourth spatial quantity, the “spissitude” that is added to length, breadth, and height. In More’s world, the whole spiritual life occurs in the fourth spatial dimension. Objects have or do not have spissitude, which means they lack or display a spiritual dimension.[253] Later on, Hinton added to the cardinal directions up/down, north/south, east/west the directions in spissitude: *ana* for increasing spissitude and *kata* for decreasing it. From More’s Neoplatonism one can derive a plethora of theories about what (or who) is populating the hyperspace. In contrast, Kaluza and Klein’s ideas to put some special fields onto the x^4 seem very modest compared to what were the inhabitants of the extra dimensions before and during the 19th century. No wonder that extra dimensions have had one of the worst scientific reputations in the nineteenth century: ectoplasm, vital forces, consciousness, light, radioactivity, angels, spirits, and even God

[253] It is an interesting discussion whether More was a materialist or an idealist in this respect. See (Henry 1986, 172-195). Adding a spatial dimension to account for spirituality is a form of materialism from my point of view.

were inhabitants of the extra dimensions of the physical space. In other circles, extra dimensions were associated to thaumaturgy and mysticism.[254]

In *Transcendental Physics (1878)*, the astronomer J. Zöllner tried to show that spirits and ghosts lived in the fourth dimensions. More seriously, he hypothesized that Newton's action-at-distance can be viewed as an action by contact in 4-D. Riemann speculated that light is a manifestation of the fourth dimension and that ethereal bodies live in higher dimensions. This speculation is maybe the first attempt to associate electromagnetism to extra spatial dimensions. Early sci-fi literature populated the extra dimensions with villains who could rob any vault or get into any room whatsoever (The 4-D Auto-Parker by B. Olsen 1934):

A crook could pilfer bonds and stocks,
Then laugh at prison bars and locks;
One step in this direction queer,
And presto! He would disappear!

Needless to say that everything that happens in D>4 is more gruesome and terrifying than our normal D=3 world. It seems that once we accept that some access to extra dimension is possible, all hell's breaks loose, literally, as in Plato's myth of Gyges. Also the morality or even the concept of life becomes relative to the dimensionality. In the novel "Hellhounds of the Cosmos", C. Simak for example thought of life in higher dimensions as

[254] P. D. Ouspensky attached to *The Model of the Universe* (1922) a chapter on the fourth dimension. He claimed that snails have a 1-D consciousness, ordinary humans have a 3-D consciousness and mystics strive to attain a… 4-D consciousness. Christian theology also explored the very idea of extra dimensions in order to provide explanations for miracles. Seemingly Abbott's book can be interpreted in a Christian key in which the Sphere is a messenger from God, similar to angel Gabriel. Also, more recently "hyperspace" was interpreted again within Christian theology in (Hudson 2005, 223) especially in respect of the mysteries of the Christianity. I do not endorse these types of interpretation and I prefer to keep my analysis of extra dimensions as physical as possible as and less metaphysical than these authors.

being superior to life in 4D. For the D>4 monsters, "life which is one dimension above us in evolution", we are nothing more than vegetables (Nahin 1993 86). In fact, what seems ugly and incomprehensive in D=3 can look beautiful and round in D=4. Ditto about moral and ethical matters.[255]

This was a serious business of some of the top illusionists and even the mathematician Felix Klein wrote against Zöllner's claims. Lord Thompson was also involved in judging Slade. More recently, souls, angels, demons, villains, superheroes, have lived in extra dimensions in Sci-Fi novels.

Epistemic dimension. It would unfair not to mention here the romanticized approach to extra-dimensions taken by Abbott (*Flatland,* 1884) who popularized deep and complex non-Euclidian geometrical concepts to the large public.[256] In Flatland (D=2), some inhabitants are 2-dimensional, some are 1-dimensional (their Kingdom is called Lineland), some are regular, some are irregular etc. Once in a millennium, a Sphere enters the Flatland and talks to Mr. A. Square, a mathematician from Flatland. It is interesting that Mr. A. Square could in principle grasp the D=3 or even the D=4 world, although he could not visualize it. He saw no limits in respect of the dimensionality of the real world. For several reasons, the Sphere does not take further this discussion.[257] One can interpret

---

[255] The idea that morality or beauty is relative to dimensionality is also explored from a theological point of view in (Hudson 2005, 223).

[256] The novel is usually interpreted as a social satire and not as a book about geometry. Low class citizens, including women are curves or lines and rich people are 2-D regular shapes. Color was also the attribute of emancipated classes and some liberals tried to impose it but the revolution of colors was put down by the establishment. Abbott was able to combine interesting Victorian mores with very deep and mathematical concepts. See (Rucker 1977, 133) for a mathematical reconstruction of *The Flatland*.

[257] *Sphere*: […] But men are divided in opinion as to the facts. And even granting the facts, they explain them in different ways. And in any case, however great may be the number of different explanations, no one has adopted or suggested the theory of a Fourth Dimension. Therefore, pray have done with this trifling, and let us return to business.

Abbott's book as an argument for the logical possibility of a world with D>3, whichever this D would be. But even after Abbott's incredible success, the mysticism surrounding the higher dimensions prevailed and it constituted a strong reason to shun away from any attempt to provide a scientific argument. There is another epistemic component in Abbott's novel: it seems that we cannot visualize extra dimensions, but we can conceive them (as Mr. A. Square was willing to). A later piece of science-fiction that discusses the wormholes in within such a closed spherical world is (Burger 1983; 1965, 208). The wormholes are usually gates to other parts of the ordinary space, but they act as shortcuts.

Physical access to hyperspace. In the 20th century, artists and clairvoyants have tried to depict or to access extra dimensions. Some tried to make a living from extra spatial dimensions. Among others, Henry Slade and Johann Zöllner claimed that they were able

---

I (Mr. A. Square=the narrator). I was certain of it. I was certain that my anticipations would be fulfilled. And now have patience with me and answer me yet one more question, best of Teachers! Those who have thus appeared - no one knows whence - and have returned - no one knows whither - have they also contracted their sections and vanished somehow into that more Spacious Space, whither I now entreat you to conduct me?
Sphere. (moodily). They have vanished, certainly - if they ever appeared. But most people say that these visions arose from the thought - you will not understand me - from the brain; from the perturbed angularity of the Seer.
I. Say they so? Oh, believe them not. Or if it indeed be so, that this other Space is really **Thoughtland**, then take me to that blessed Region where I in Thought shall see the insides of all solid things. There, before my ravished eye, a Cube, moving in some altogether new direction, but strictly according to Analogy, so as to make every particle of his interior pass through a new kind of Space, with a wake of its own - shall create a still more perfect perfection than himself, with sixteen terminal Extrasolid angles, and Eight solid Cubes for his Perimeter. And once there, shall we stay our upward course? In that blessed region of **Four Dimensions**, shall we linger on the threshold of the **Fifth**, and not enter therein? Ah, no! Let us rather resolve that our ambition shall soar with our corporal ascent. Then, yielding to our intellectual onset, the gates of the **Sixth Dimension** shall fly open; after that a **Seventh**, and then an **Eighth**—
How long I should have continued I know not. In vain did the Sphere, in his voice of thunder, reiterate his command of silence, and threaten me with the direst penalties if I persisted. Nothing could stem the flood of my ecstatic aspirations. Perhaps I was to blame; but indeed I was intoxicated with the recent draughts of Truth to which he himself had introduced me. However, the end was not long in coming. My words were cut short by a crash outside, and a simultaneous crash inside me, which impelled me through space with a velocity that precluded speech. Down! down! down! I was rapidly descending; and I knew that return to Flatland was my doom. One glimpse, one last and never-to-be-forgotten glimpse I had of that dull level wilderness—which was now to become my Universe again—spread out before my eye. Then a darkness. Then a final, all- consummating thunderpeal; and, when I came to myself, I was once more a common creeping Square, in my Study at home, listening to the Peace-Cry of my approaching Wife.

to untie a knot whose ends were sealed, get out of a box without going through the sides, or walk into a box without entering it.

Asimov's *Little Lost Robot* (1947) depicts a hyperdrive and hyperspace is used as a shortcut for travels faster than light. It most frequently involves a gate. In the *Foundation's Edge* hyperspace is a condition. There is no velocity in hyperspace, but seen from 4-D, the velocity looks infinite. Mass in real space distorts hyperspace. Hyperspace is accessible only from empty real space. In other depictions, space is folded by a technology (*Dune*), hyperspace is the subspace or an inner space (*Star Trek*) and space is like a wave can be ridden (*Space Battleship Yamato*).

The literature on the possible access we have to extra dimensions of spacetime is huge. Besides the speculations inspired by recent String Theory involving multiverses and string landscape, in the last decade one can see a resurgence of literature on extra spatial dimensions in cognitive sciences, literature, art, esthetics, theology, etc. Whether our brain is wired to reason about objects in D>3 or not, even if our visual system is not able to represent them, is an interesting topic and it was debated since Von Foerster's experiments from 1970s in which subjects were able to perform "legal" D=4 maneuvers, spot inconsistencies, perform movements and tasks, etc.[258] Maybe Abbott had a point in depicting Mr. A. Square as being able to *reason* about D>2, albeit he was not able to *visualize* it. Th. Banchoff used in the 1980s colors to help us visualize hypercubes. What is interesting is that he made the clear distinction between slicing and projections because these two strategies are the main ways of helping us visualize hyper-objects. Recent works on

[258] These reports were never published but they are available in the Biological Computer Laboratory at UI, Champaign reports no. 712 and 722. (Banchoff 1990, 210) is based on this early research.

visualization of D>3 are based on computer simulations. It seems that walking in 5-D or playing video games in 5-D is possible.[259] Even if the arguments that we could conceive hyperobjects are sound, they are not relevant for the present discussion. Remember that in Kaluza there are no fields with variation on the fifth axis. In Klein, the fifth axis is compactified with a tiny radius. Objects would look different in the two 5-D manifolds. Because of the cylindricity condition **CYL**, Kaluza's objects are extended in the fifth dimensions. Klein's objects are not visible because they of the compactification **COMP**. In other words, the best guide in reasoning about Kaluza-Klein extra spatial object is mathematics and not our intuition or our hypespace reasoning because of the special topology these spaces have.

Making the difference between science and pseudo-science of extra dimensions, between relevant arguments and irrelevant arguments, is not always easy. One reason among others is that once one postulates extra dimensions, almost anything that happens in 4-D can be explained in D>4 dimensions. In almost all cases pseudo-explanations tend to overcome scientific explanations. Critics of String Theory or of the Kaluza-Klein theory have the same feeling that these extra-dimensional theories display largely inflated explanations and pseudo-explanations. Sometimes and somewhere down the road the science of spacetime tends to become more mystical and obscure than one can take. Is it important to discuss this hyper-inflated explanatory power of extra-dimensional theories in the context of Kaluza-Klein theory. Overstating the explanatory power of any theory that brings in

---

[259] A good reference for the projective technique in art is (Robbin 2006, 137). Recently, there are several computer games which claim they are walking the players in extra dimensions.

extra dimensions of spacetime is evident and I will come back to this issue especially in the context of Klein's theory.

The mathematics of hyperspace. What is in fact the mathematical idea of hyperspace? One can think that some (one) extra spatial dimension has something special in respect of the other three known spatial dimensions. A very respectable mathematical idea, the non-Euclidian geometry, was based on the extra dimensions of mathematical spaces, having different topologies and metrics than the Euclidean space. The very fact that one of the dimensions can be too small (or too big) to be perceived has been speculated based on a physical theory by C. S. Hinton as early as 1888.[260] He speculated that variations in extra dimensions are too small to be perceived. It is important to know that the fundamental motion Hinton associated to the fourth dimension were "double rotations". As far as I know, the idea of circular motion in the fourth dimension originated with Hinton:

> If four dimensions exist and we cannot perceive them, because the extension of matter is so small in the fourth dimension that all movements are withheld from direct observation except those which are three-dimensional, we should not observe these double rotations, but only the effects of them in three-dimensional movements of the type with which we are familiar.
>
> If matter in its small particles is four-dimensional, we should expect this double rotation to be a universal characteristic of the atoms and molecules, for no portion of matter is at rest. The consequences of this corpuscular motion can be perceived, but only under the form of ordinary rotation or displacement. Thus, if the theory of four dimensions is true, we have in the

---

[260] See (Hinton 1904 (1912)). His essay containing the argument that the fourth dimension was not observable was published in 1904, reprinted in 1912. In an early work (*Scientific Romances*, 1884) Hinton was absolutely fascinated by the idea that time can be associated to… the fourth dimension. Classical mechanics can be reproduced in 3-D+time. Becoming and motion in time are due to a constant motion of a plane in D=4, similar to what Minkowski suggested. 3-D symmetries are not anymore absolute: for example, left and right handiness are not absolute and any form can be transformed into its mirror image by a rotation in D=4. Links of a chain may be separated without breaking them in 4-D. Density and the four states of matter are relative to the dimensionality. Hinton prefigured special relativity in several ways, but more philosophical investigation of his early work is needed in order to decide what the conceptual differences with **SR** are. For example, Hinton has no concept of metric and signature and he did not realize that a special measure on the fourth dimension is necessary. See **(144)** below.

> corpuscles of matter a whole world of movement, which we can never study directly, but only by means of inference (Hinton 1904 (1912) 223).

The transition from the analysis of classical, massive bodies in four spatial dimensions, to the analysis of metric fields in five dimensions (four spatial and one temporal) is a major leap forward, indeed, but the idea that variation of fields and motion along the fourth dimension can be constrained to small values belongs to Hinton. He did not intuit Klein's compactification **COMP** but he had an intuition close to Kaluza's **CYL**. It is not clear to me whether G. Nördstrom, Th. Kaluza or O. Klein had access to Hinton's work in Germany or they knew about his work indirectly. It is also not clear whether Hinton's work was translated or available in Germany to Kaluza in the early 1920s.

Hinton was a mathematician without serious interests in physics. His inclination to adopt Platonism in mathematics is clear from the way in which he introduced the ideas of Parmenides and Plato in his main monograph on 4-D (Hinton 1904 (1912)). Geometers have analyzed strange topologies and n-dimensional spaces since Gauss, and there were attempts to link a non-Euclidian topology of an invisible spatial dimension to physical properties of the real world.

15.7. Dimensional explanation

Mathematicians have had a more relaxed attitude toward spatial dimensions. Geometry does not restrict itself to three dimensions and despite any lack of empirical support we could live in a spacetime with D>3. Indeed, D=3 can be taken as a prejudice: "abstract studies such as these allow one to observe relationships without being limited by narrow terms, and prevent traditional prejudices from inhibiting one's progress" (Riemann 1953 268).[261]

According to the Campbell-Maagard theorem, embedding the 4-D space in a higher dimensional space is perfectly kosher from a mathematical point of view (see 13.1 and 15.6). But this result does not tell us how explanation can work; it just opens the possibility of some physical results. In this section I discuss precisely the explanatory power of dimensional reduction.

Dimensional reduction is not a mathematical operation of truncation. All our scientific theories are highly dependent on what I call here "dimensions". In this context I am interested in dimensions of the spacetime manifold; under some caveats but in a broader context, dimension can be any degree of freedom one can encounter in the practice of science. Degrees of freedom do have explanatory power and are also endowed with relevance or scientific significance beyond being mathematical objects, even if they are not dimensions of the *physical* space. We need to know which degrees of freedom or dimensions are relevant to a specific situation. In this respect, the degrees of freedom are highly dependent on the context of our scientific inquiry or more general dependent on us as investigators. Dimensions in this general meaning are just ways of investigating the way in which systems evolve, interact with the environment, etc. We impose dimensions upon the system we analyze like we impose other conditions: idealizations, abstractions, constraints, etc.

Why a new dimension? The standard accounts of explanation briefly reviewed in Section 3.1 do not seem to fit the "dimensional explanation". The 5-D world is not more

[261] Riemann's habilitation thesis was written in 1854, but published in 1868.

familiar, it is not easier accessible than the 4-D, it does not provide foreseeable predictions or applications. Kaluza and Klein hoped that the fifth dimension helped us understanding the **GR** and **EM** in 4-D. **D-N** model seems also a little bit off in the case of dimensional reduction, too. There no way to separate the laws from facts: it is not clear whether 5-D can play exclusively the role of laws and 4-D the role of facts or observations—this can be a promising way of interpreting Kaluza and Klein as illustrating the D-N explanation but I won't pursue it in details here because the distinction between cover laws and facts is problematic in the premises. It is difficult to see what qualifies as phenomena and what a law is here. Kaluza-Klein explanation explains laws and phenomena in 4-D not only phenomena.

I prefer to take dimensional explanation to more concrete level and discuss it as geometrical explanations. They always come with an interpretation by which several geometrical features are endowed with physical significances.

Dimensional reduction is present in Kaluza and especially in Klein. Whether dimensional reduction is the right term to be used in Klein or even in Kaluza is a relevant question. Almost all authors ponder Kaluza-Klein theory as being premature and underdeveloped (Blagojevic 2002 334). The fundamental interactions were not understood yet, the spontaneous symmetry breaking idea should wait another five decades to be discovered, **QM** was understood as wave-particle mechanics etc. Since the 1930s, the geometrical explanation fell into disrepute mainly because geometry was not able to provide explanations and models of matter, at least less successful than the **QM** explanations. For decades, speculations about curled-up extra dimensions seemed to Smolin "as crazy and unproductive as studying UFOs. There were no implications for experiment, no new predictions, so, in a period when theory developed hand in hand with experiment, no reason to pay attention" (Smolin 2006 52).

Smolin also points out that playing with dimensions is a shaky game:

> The more dimensions, the more degrees of freedom — and the more freedom is accorded to the geometry of the extra dimensions to wander away from the rigid geometry needed to reproduce the forces known in our three-dimensional world (Smolin 2006 51).

This inflation of models chases nowadays' String Theory, too. The supersymmetric theories are so rich that they can explain almost any imaginable universe. And this affects Kaluza-Klein generalizations which seem to be nothing more than a mathematical tool of representation and not a physical theory that reflects reality.

This reflects in fact an older practice of physics. For many physicists, the main activities to be performed in physics between 1930s and 1970s were (a) accurate calculations and (b) providing better predictions to measurable phenomena. I claim that from a foundational point of view this is incomplete. After the 1960s, geometrical explanations have gained a slow impetus especially by the development of the fiber bundle formalism. Later on, a geometrical formalism was developed even for **QM**. Both Kaluza and Klein tried to look further than calculations and accurate predictions toward some non-empirical virtues of their theories.

Recent developments in dimensional explanation. The fate of the dimensional explanation is similar: well beyond the context of Kaluza-Klein theories we do not know whether dimensional explanations are right or wrong. Kaluza-Klein theory eventually resurfaced due to its generalization to Yang-Mills fields by adding extra dimensions with more and more sophisticated topologies and by including quantum effects—perhaps the milestone of its development is Witten's articles from 1981 (Witten 1981b; Witten 1981a). Witten and others contributed to the resuscitation of the Kaluza-Klein theories in the 1980s, but they were agnostic about them being true or false: "What we do not know is whether the time is finally ripe for the Kaluza-Klein theory, whether there still are crucial things we do not know, or whether the idea is completely wrong. Time will tell" (Witten in Capri and Kamal 1983)

For the present purposes, I am not interested in the direct observational consequences of extra spatial dimensions, compactified or not. We still have no idea what is the nature of the real spacetime at small scales: it can be discrete, it can be multidimensional or perhaps the concept of spacetime as we conceive it is meaningless. It is easy to see that at lower energies there are little or no consequences of a hypothesis such as **COMP**. At higher energy levels or at very early stages of the evolution of the universe one can come up with a different scenario: all dimensions were at the same scale, i.e. non-compactified and later on at a precise moment some dimensions changed. This specific scenario was discussed only in the last decades and it is still not well understood and underdeveloped (Bailin and Love 1987).

If empirical adequacy is so difficult to achieve, we need to look for other virtues of the theories that postulate extra spatial dimensions. In the case of String Theory unification is the most obvious sought for virtue. For the ironclad unificationists within the String community, consistency is more important than empirical adequacy. In an early advertisement of String Theory, M. Green one of the architects of the Superstring Theory wrote (Green 1986):

> [...] the unification of the forces is accomplished in a way determined almost uniquely by the logical requirement that the theory be internally consistent […] Much of the interest in superstring theories follows from the rich structure that results by requiring the theory to be consistent. [...] The fact that the quantum consistency of a theory including gravity leads to […] unifying symmetry group was an exciting development. It has led to the current wave of enthusiasm for superstring theory.

If I understand Green's argument correctly, we should aim specifically in theoretical physics to consistency, beauty, elegance, and of course, unification. The second attitude of physicists of other physicists such as Smolin and Feynman is to aim to a rather different target: chase the empirical testability of theories and eventually novel predictions. None of these attitudes are satisfactory from a philosophical point of view. Contra Green, the history of science taught us that not any piece of knowledge is a scientific theory—albeit consistent, elegant or unified. Contra Smolin & Feynman, not any well confirmed or testable theory is scientific. Both attitudes are short-sighted. It is easy to see that the former attitude of Green *et al.* brings science close to art, mysticism or religion. If you ask a physicist like Green why analyzing extra dimensions, she would reply: "Because we do not know if they are there".[262] We are in fact in a position similar to Mr. A Sphere's. But there is also the opposite attitude, transparent in Smolin's quote, which seems to deplete science of a precise aim toward knowledge: the pure empiricist wants to replicate the world in the science without paying enough attention to the gain in knowledge. Charting scientific theories of extra spatial dimensions is in fact a gain in knowledge and reflects I think the spirit of scientific enterprise because dimensional reduction and dimensional explanation can illustrate how explanation works in the case of spacetime structures enriched with extra dimensions.

[262] When George Mallory, one of the first who tried to climb on Mount Everest, was asked why he wanted to climb it, he replied: "Because it is there".

Is playing with dimensions a fruitful game in science? I argue here that in a very general context, this can bring about explanations. One can explain phenomena or explain away appearances. We looked at the sky and saw very complicated motions of planets. Then we invent epicycles. Later on we realized that the complicated motion is in fact the projection of a simple and regular motion in three dimensions. We explained away a complex appearance by using a geometrical argument. Adding dimensions can explain away appearances, too.

15.8. "Dimensionally challenged" models (dimensional truncation)

In this section I plan to start from very simple models in which dimensions are added or slashed from our representation of the world. Spacetime is a specific case of such operations, but before discussing it I look at these procedures in a broader context. I differentiate here two types of theories and their representations of extra dimensions. First we have the dimensional truncation in representing a system: be it physical, social or economical. I call them dimensionally challenged models. They use a procedure similar to Kaluza's. Second, in what I call "bare models", several conclusions can be drawn from minimal assumptions about the **5-D** world. Applied to spacetime theories, this bare model is similar to Klein's theory.

Adding or reducing the number of variables in representing a system is a computational method used virtually in any science. Dimensional reduction works with any type of variables. In some specific context, the operation of reducing the number of *variables* is explanatory. In this section I argue briefly that "dimensional reduction" is explanatory when variables are involved, not necessary spacetime variables. In all sciences, complex data come with several independent variables. In several branches of science we reduce the number of variables used to describe a system: this is an operation of truncating the representational space. A method proposed by K. Pearson in 1901 was to locally embed this large set of data in linear maps: lines, planes etc. “In many physical, statistical, and biological investigations it is desirable to represent a system of points in plane, three, or higher dimensioned space by the ‘best-fitting’ straight line or plane” (Pearson 1901, 559-572). In statistics, the line that embed the data is chosen based on a minimization procedure, the most popular being the “least squares approach”. We postulate a linear connection between variables such that some variable is independent and all the other are dependent on it. The result is a straight line or a plane. A different geometrical embedding results from a different separation between independent and dependent variable. This method is usually called the “Principal Component Analysis” in statistics and it evolved in the last decades in a geometrical direction.

Obviously enough, postulating a non-linear dependency among variables is possible. Approximation to other objects than planes is possible, too; the least square method is replaced with a different algorithm, called the K-algorithm. In sum, the component analysis allows substituting a high-dimensional vector by its projection on a best fitted lower dimensional linear manifold, where “best” means any method fitted to the concrete problem.

CM *and dimensional reduction.* Physicists want to reduce the dimensions in which they represent the world for obvious reasons: simplification, idealization, computational reasons and last but not least explanation. Fewer dimensions mean fewer variables, fewer degree of freedom, fewer equations, fewer boundary conditions, etc. We idealize motion by neglecting dimensions or degrees of freedom that are not relevant. Although a system exists in three dimensions of space, its properties display a behavior as if the system were a lower-dimensional one. For a system with *n* variables, there are several numerical methods used to reduce the dimensions of the space of analysis. A very popular one is called “feature extraction” which is in principle a mapping of the multidimensional space into a space of fewer dimensions based on a hypothesis of dependency similar to Pearson’s. This means that the original feature space is transformed by applying a linear transformation via a principal components analysis. This procedure can be used when the degrees of freedom are directly associated to the physical space.

Truncation or reducing the dimensionality of a representation is an operation that can be done in the case of simple system in physical space. In mechanics this is a no-brainer for several systems. A pendulum has roughly one spatial degree of freedom which is conveniently described by one parameter in polar coordinates (typically the angle with the vertical line). We know how to idealize planet orbits. Although it is clear that the Earth is a three-dimensional body with an orbit in 3-D almost all models of its orbit around the Sun are two-dimensional. If we choose the distance between Earth and Sun as one parameter and the angle as the second degree of freedom, we can eliminate from all equations the third angle. The revolution of Earth occurs in a plane called the “ecliptic plane” and we reduce our representation to that plane, mainly for computational reasons.[263]

---

[263] In our Solar system, not all planets are in the same ecliptic plane, although the differences are minor. Earth has an extra degree of freedom because of its rotational axis is tilted 23.5° from the ecliptic plane. There is a slow change in time, but it is negligible. The ecliptic plane changes very slowly during the history of the Solar system because of astronomical perturbations, solar winds, asteroids, drag effects, etc. Similarly, galaxies have an ecliptic plane which surprisingly is very close to the planetary ecliptic plane(s).

The idea of this example is that the third dimension, although present, can be neglected for many practical purposes.[264] But there is a tricky part easy to overlook. We cannot completely ignore physical quantities living on the z axis. The most important quantity that characterizes the motion of the Earth around the Sun, its angular momentum, is a vector that is parallel or very close to the z axis. Metaphorically speaking, the angular momentum lives on the z axis although it characterizes a system that lives entirely in the (x,y) plane. If we take into consideration Earth's rotation, the angular momentum has components on all axes (although the dominant one is on the z axis). The strategy of completely ignoring the z axis, even for such simplified system, has few chances to succeed. Useless to say that the angular momentum has important explanatory powers and its conservation is not at all a trivial fact. We can think that maybe something more specific can be said about the z axis, for example that all vectors have constant components (in time) along the z axis. The moral is that whatever happens in the *(x,y)* plane does not stay there. The Flatlander does not see the *z* components of some vectors, although they explain her evolution is *(x,y)*. She feels the effect of the conservation or on the contrary of the variation of L_z, for example. Lots of gruesome appearances in 2-D can be explained away by a simple, 3-D model.

A simple, "dimensionally challenged" ***EM***. In other theories than **CM**, dimensional truncation can have even more dramatic consequences. In **EM** the intricacy of completely ignoring one dimension is even more manifest. Is there a way to reduce **EM** to two spatial

[264] But this is a contingent feature of our solar system. In other cases, other solar systems can have extra-degrees of freedom and a much quicker motion in the third dimension of their orbits. Of course a further investigation is to explain why the Earth evolves in one plane only. I suspect that a combination of initial conditions and the stability of our solar system explains why the Earth's orbit is more or less two-dimensional. Of course it is a contingent fact that here in this Solar system all planets have an ecliptic plane. Only higher order effects on the motion of Earth and it is not significant in normal conditions.

dimensions? Like before, a quick answer would be to drop from all equations the dependency of the z axis. Because of the Lorentz force and because of presence of curls in Maxwell's equations, cutting a dimension is a bad idea. For a movement of particles in the (x,y) plane, magnetic vectors will be always perpendicular on that plane. In other words, magnetic fields are not present in the (x,y) plane although they explain the dynamics of charged particles moving in that plane. One suggestion is that in this case, clearer than in the model of the planetary system, we *cannot* ignore the z direction. How do we reduce the dimensions? One can take the (x,y) plane and postulate that the dynamics occurs only in that plane by ignoring the dependency on z; in other words we ignore the time derivative of all quantities on the z axis. This is not the same as saying we ignore the existence of the z axis. It does not mean we cut the z axis, it is just a dynamical hypothesis:

[89] The 3-D (i.e. D=3) **EM** theory requires that all quantities should have no z-dependence.

The expression of the Lorentz force can help us dealing with such a dimensionally reduced system. The variation in time of the momentum is:

$$\dot{\mathbf{p}} = q\left(E + \frac{\mathbf{v}}{c} \times \mathbf{B}\right) \tag{151}$$

If the particles lives in the (x,y) and the magnetic field is only in the same plane, the magnetic force will be exerted in the z direction and it will pull the particle from the (x,y) plane. We cannot drop the z components of the vectors! But in order to keep the particle in the (x,y) plane, we need a magnetic field on the z axis and null fields in the z direction. The electromagnetic field with: $E_z=B_x=B_y=0$ will indeed keep the particle in the (x,y). The problem is this: in the (x,y) plane B_z is not a vector component, but acts as a scalar field. So in D=2 we can talk about electromagnetism only if we add a scalar field B_z to a vector field $\vec{\mathbf{E}}(x, y)$ in the (x,y) plane (Zwiebach 2004 41-42). We do not ignore the z axis and we do not think as all quantities being zero on the z axis. The weaker condition is that we require that there is no dependency on the z coordinate. Sounds familiar? This is similar to Kaluza's **CYL** condition.

The lesson here is that here, too, components on other dimensions are explanatory: we can ignore them in our predictions or calculation but they may nevertheless have explanatory power. The main problem is to decide whether it makes explanatory sense to go beyond the usual dimensions of physical space.

Are vectors involved in Flatlander's **EM** theory? Some are derived from our 4-D **EM** theory. The Flatlander has a strategy to solve this problem. She may want to add a scalar field which looks arbitrary in D=2 and indeed is arbitrary if the Flatlander insists she lives in two dimensions. She needs to postulate the extra scalar in D=2 in order to explain some forces she experience, for example the Lorentz force when her test particle hits an area with magnetic field. But in fact this scalar is related to a three-vector which does not have components in (x,y). The B_z is a brute field or a brute fact for the Flatlander, although for the observer in higher dimension it is not brute, but a component of the **B**. Needless to say, B is related to the field **E**, all being related one to the other by Maxwell equations, the one that has components in Flatland. Simply put, what is a arbitrary or ad-hoc brute fact for the Flatlander is in fact not a brute fact and definitively not an ad-hoc assumption for the D=3 observer. In other words, some of these toy examples illustrate that what is brute and what is ad-hoc, as well as explanations and predictions are somehow indexicalized to the number of dimensions we take into account.[265]

*The "dimensionally challenged" **GR***. Let us go back to the context of spacetime theories. How does look a serious **GR** in Flatland, even a Lineland**?** The procedure of dimensional reduction, inspired by projective geometry, can be applied to anything, physical systems included: in fact, space coordinates are variables and can be conventional how we relate them to the dimensions of spacetime. Unlike **EM** or mechanics in D=2, gravitation in D=2 tells us a sum of things we are interested in. It is not a simple toy model and it was studied in the 1980s by S. Deser, R. Jackiw and Templeton. Interestingly enough, 3-D gravitation illustrates Mach's dream better than 4-D gravitation: in D=3 there are no gravitational excitations and matter determines locally and entirely the geometry (Deser 2003 397). Also, the Riemann tensor equals the stress tensor. A very intriguing possibility, although totally unrealistic, D=2 (i.e. 3-D) gravity has some amazing features that are useful in understanding better the real 4-D **GR**. As expected, fewer fields can be described in such a dimensionally challenged **GR**. The dimensionally challenged **GR** shows us what role truncation plays when spacetime degrees of freedom are involved.

15.9. A "bare model" of the 5-D spacetime

The fears that unification can become a scientific aim in itself can be addressed by building a bare model which is not unificatory, albeit explanatory. I started this chapter with the skeptical attitude according to which spacetime in itself explains nothing. As a counterargument, I build here some simple models in which the properties of spacetime

[265] A somehow more convoluted example can be given for quantum systems where dimensionality plays a crucial role. I prefer to keep my analysis classical and stay closer to the original model of Kaluza and Klein.

explain without being unificatory. I also address here the concern expressed by [56] (p. 263). What if all what Klein achieves is nothing more than mere results or consequences of a unificatory hypothesis and not explanations in themselves.[266] Although neither Kaluza nor Klein proposes such a fictional construction, they can be easily constructed having in mind similar assumptions. Explanations which are strongly related to unification are consequences of unification and they can be suspected as not being explanatory: mostly Kaluza's results are in this category. In the following chapter I show that Klein's results are not all consequences of the unificatory assumptions and some qualify as explanations independently of unificatory assumptions.

Let us take here a simple model in which we assume a peculiar topology of spacetime, i.e. compactification of a spatial dimension, and see whether it can explain or not phenomena in 4-D. A bare model is not unificatory and adds the minimal structure to the **5-D** spacetime and has minimal assumptions regarding the topology of extra dimensions. The bare model is not unificatory, albeit it explains and explains away phenomena in 4-D. A procedure to decide whether a structure S can explain is to create a model containing S, put it in an "empty universe", and see what consequences can be drawn from this model. Sometimes the structure itself can provide information about the theory we are trying to investigate. Highly idealized models, albeit totally unrealistic, are explanatory powerful and in fact can help us in understanding more complicated theories. I mentioned briefly that **GR** in D=2, i.e. gravitation in two spatial dimension and time, is a gem of interesting results. Equally important is to look at spacetimes with D=4 (5-D) spacetimes with different topologies and characteristics, mainly their symmetries. The D=4 models and the

[266] Thanks to Nancy Cartwright and to Christopher Smeenk for stressing this aspect of my argument.

D=2 are both relevant albeit in two different ways. In the spirit of the Campbell-Magaard theorem, as D=4 bare models are depleted of physical fields we look for their explanatory power for thing observable in D=3. In D=2 we postulate special types of field in order to describe gravity in D=3.

Are spacetime symmetries without any field important in themselves? Let us take a look at a model in which there is nothing on the 5-D spacetime, except its Riemannian structure. Similar to Klein's analysis, I look here at manifold with a given symmetry, deplete it of any fields. What this structure has is nothing but its symmetry and the given Riemannian metric.[267] Let us simplify and for notational purposes take the "special axis" $x^4=z$. We can adopt the **COMP** hypothesis of Klein here and let z take values in the interval $[0, 2\pi\lambda_4]$. From the beginning it is clear that the winding number, the modulus induces and ambiguity called the "moduli problem".[268]

The vacuum of our universe is given by:

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu - \left(\frac{R_4}{\lambda_4}\right)^2 dz^2 \tag{151}$$

The only entity allowed is a massless, spinless probe particle that moves on geodesics. They are only probe particle: they do not create wrinkles in the structure of the spacetime. Of course, this is already a gross With Klein, I assume that a massless particle

[267] The present analysis was carried by in early 1980s by Witten i.a. My exposition follows partially (Blagojevic 2002, 522).

[268] The winding number is somehow similar to the reading of a watch. I look at a clock and I see the arm pointing at 10:25. The angles tell me something about the time elapsed since, let us say, 12:00am. The winding number can be equivalent to the "day" reading, i.e. Thursday. It is not directly present on the clock, but it is clear that the hour arm moved to that position by crossing *n* times that position in previous days. If we read only "hours" on the clock, the day information is not relevant although it is present in the calculation. The winding number constitutes a major difference between the Kaluza-Klein theory and String Theory: the strings are characterized by their winding around a compactified dimension, whereas in Klein the winding number does not have a physical significance. This makes sense if you think that the tension of a rubber band winded around a cylinder depends on the "winding number".

with no spin will follow a path that satisfies the conservation of the P, i.e. the momentum-energy tensor:

$$P^m P_m = 0 \tag{151}$$

After some algebra, the final equation of motion of the probe particle is given by:

$$p^\mu p_\mu = p^4 \left(\frac{R_4}{\lambda_4} \right)^2 \tag{151}$$

which suggests that a mass of:

$$M = \left| p^4 \right| \frac{R_4}{\lambda_4} \tag{151}$$

can be associated in 4-D to this massless propagator in 5-D. A first conclusion is that the vacuum solution in 5-D can create the “illusion” of massive particles propagating in 4-D as if they have mass. The bare model teaches us an important lesson: matter can be taken to be an independent feature of reality or on the contrary it can be related to structures of the D>5 spacetime (here, the *compactified* extra spatial dimension creates masses). It also helps us identifying the explanatory components of the Campbell-Maagard theorem.

In the bare model there were no fields at all, only a given symmetry and the conservation law that is “correlated” to this symmetry by Noether’s theorem. Mass in 4-D, here the much-speculated “graviton” can be generated by a compactification of x^4. The bare model “explains away” the illusion of masses in 4-D which are explained away by the structure in 5-D.[269]

[269] The business of explaining away some of our illusions is commonly encountered in science. The heliocentric model explains away the illusion we have that the Sun is moving. The Everettian interpretation of quantum mechanics explains away the illusion that there is only one world, that there are determinate outcomes of measurements, etc.

The bare model bears similar conclusion to Klein's, although it is weaker. Adding more fields to the 5-D structure enriches its explanatory store and consequently explains away more illusions in **4-D**. When a specific type of g_{mn} field is placed on the **5-D** manifold, one can explain the illusion that there are charged particles and that there are photons, as we will shortly see. In this section, the result was simpler and more intriguing: a massless field in **5-D** can create the illusion of a massive particle in **5-D**.

This is not surprising at all if we think that an object that has a natural trajectory in **3-D** let us say a circular trajectory, appears as growing, shrinking, appearing or disappearing to an observer in a Flatland **2-D**. For more complicated objects in **5-D** such as branes, the consequences of such projections are even more dramatic: even gravitation itself can be generated from the motion of branes (Randall 2005 450).

In this chapter I showed that there is room for both explanations and pseudo-explanations in spacetime theories. Kaluza's theory is similar to a dimensionally challenged theory in **5-D** where fields are considered smooth on x^4. On the contrary, the bare model in **5-D** is closer to Klein's theory who hypothesized that the fifth dimension is compactified.

In the following chapter I focus on the explanatory power of Kaluza's and Klein's theories and their relative explanatory store.

Chapter 16. Dimensional explanation in Kaluza-Klein theories

In the previous chapters I separately discussed unification and explanation in several contexts. The aim of this chapter is to mesh back unification and explanation. But I do not want to connect them too strongly as I avoid to envisage explanations as mere consequences of the unification procedure. I envisage Kaluza's and especially Klein's theories as illustrating a two-stage process. First the extra dimension is assumed; this is the unificatory stage that has been discussed in Chapter 13 and Chapter 14. The second stage is explanatory in nature and it is based on the "dimensional reduction" where the explanation of the **4-D** world from the **5-D** theory is sought. I call the second, explanatory stage, the "dimensional explanation"; the word "reduction" can be misleading in this context, especially as philosophical term. In a broader context, dimensional *explanation* acts as predictive mechanism or as problem solver and operates independently of unification: in the dimensional explanation, aspects of low dimensional physical world are explained as "projections" of a higher-dimensional world. The idea of projecting properties from a richer structure to the **3-D** world is frequently used in quantum mechanics where the ultimate ontology is a wave function living in a configuration space.[270] Ultimately, in Kaluza-Klein theories the classical field g_{mn} lives in **5-D** and its projection onto a **4-D** manifold has or is supposed to have explanatory role. There is a variety of procedures of projecting a theory in D dimensions down to D-1, D-2 etc dimensions. Geometry helps here a little: the projective geometry was a well established topic of research in the 19th century. Bodies, curves, relation between objects, trajectories, virtually any structure can be pro-

[270] Not all interpretations of quantum mechanics agree on this claim. It seems that for some authors the primitive ontology does not include the N dimensional wave function at all.

jected down to a lower dimensional subspace. This operation of projection is explanatory, too. Think of the classical example of the shadow if the flagpole as a dimensional explanation. We explain the shape and the length of the shadow by the more challenging question is to project down laws and equation from 5-D onto 4-D. The main question is (a) whether we can explain phenomena happening in 4-D and (b) how many phenomena can be explained as such.

In the context of Kaluza and Klein, dimensional reduction is mainly explanatory and problem solving. Unlike the "dimensionally challenged" and the "bare" models discussed in Sections 15.8 and 15.9, in both Kaluza and in Klein unification and explanation are intertwined. The theory also acts as a problem solver and a problem maker, as it were. Unification is achieved by adding a new structure to the existing spacetime, structure that now *unifies* **EM** and **GR**. At a second stage, from the new, empowered **5-D** structure, Kaluza and especially Klein tried to *explain* as much as possible of the **4-D** world. The intended *explananda* are the **4-D** laws of **EM** and **GR**. Beside the laws of **EM** and **GR** in **4-D**, in Klein's case there are other, unintended *explananda*: the quantization of charge, the symmetry of **EM**, as well as other more remote explanations: the existence of positrons, the existence of graviton, the magnetic monopole among other things. I show that the second stage is as important as the first one: both authors tried to explain laws and phenomena in the four-dimensional world as aspects or projections of laws in the **5-D** world.

There is a possible problem with this two-stage process. It seems that there is an internal conflict between adding mathematical structures and simplifying the apparatus. In Kaluza-Klein I see the dimensional reduction as operating after unification has been achieved. In the two-stage process, explanation and unification are coupled in at least two senses: (a) explanation comes after unification and (b) unification is designed having explanation as a purported aim. On the other hand, unexpected and novel explanations are present in Klein's theory. Both Kaluza and Klein added structure and then would show that **EM** and **GR** are in fact simpler when unified. These are two conflicting features of both Kaluza's and Klein's theories. On the one hand, there is a qualitative parsimony, at least in Klein's case, because he used fewer types of symmetries and fewer constants than **EM** and **GR**. In what sense are their theories more parsimonious than the conjunction of **EM** and **GR (EM&GR)**? I think they both illustrate a form of qualitative parsimony discussed in the philosophical literature. What I mean here by "qualitative parsimony" is roughly the number of types of entities postulated by the theory, suggested in Friedman's analysis and discussed more recently in the literature (Baker 2003 247; Nolan 1997, 329-343). Although "parsimony" is highly sensitive to the language in which the theory is formulated I claim that the differences in language do not play a preeminent role in comparing the parsimony in Kaluza and Klein and consequently Klein's theory is more parsimonious compared to **GR**&**EM**.[271] In Kaluza, there are some hypotheses which seem at least *ad-hoc*, mainly the **CYL** condition. One can see that if there was parsimony in Kaluza, it came with the high price of *ad-hocness*.

On the other hand, Kaluza and Klein both added geometrical structures to the Riemannian space $\mathbb{V}^4$ and this runs against the aforementioned qualitative parsimony. They also needed to postulate new fields and new constants. Both Kaluza and Klein struggled to minimize the unwanted effects of the new structures by making some as-

[271] In other cases, parsimony is highly problematic precisely because a different formalism with the same consequence can look more parsimonius.

sumptions, especially in relation to the field ϕ. Despite the successful unification achieved, there was a surplus structure to be accounted for. By enlarging the geometry of their theory, Kaluza and Klein counterbalanced the gain in relative parsimony of their theories. Kitcher suggested that a discussion of sheer numbers of brute facts is not well founded and needs several qualifications. As suggested in Part I, Friedman's definition of unification is not satisfactory either, per Kitcher. I do not plan to count the *number* of brute facts of these two theories, but a discussion of *types* of brute facts present in both is germane here.

Adding structure is not in itself valuable if it is not followed by a clear gain. At the end of my analysis it will be clearer that Klein's intended gain was not in elegance, simplicity or parsimony; he aspired to greater knowledge of the world. What does the fifth dimension teach us about the world? Arguably, for Klein it explains the world better than **EM**+**GR** did. Bringing explanation in is a very important step one need to take when one estimates non-empirical virtues of theories. I argue that Kaluza and Klein ascertained explanation, more precisely "dimensional explanation". But again I will adopt a comparative method and show that Klein's explanation has a larger explanatory store than Kaluza because there are several major differences in Klein compared to Kaluza in the way unification and explanation are coupled.

In the Kaluza-Klein context, dimensions are explanatory. Later on, dimensions would eventually provide some predictions. In the further development of the theory, dimensional explanation came after unification. This chapter eventually goes beyond the historical episode of Kaluza and Klein. In order to argue for the explanatory power of the Kaluza-Klein theory, we need to reconstruct and reinterpret their theory in several different, but akin contexts.

Table 2: The dimensional explanation in Kaluza and Klein

	4-D structure	**5-D** structure
Step 1: Unification	1 extra dimension + identifications →	
Step 2: Explanation	← Dimensional explanations & interpretations of 4-D entities	
What lives in this structure?	$g_{\mu\nu}$, $F_{\mu\nu}$	*Kaluza:* g_{mn}, ϕ *Klein:* g_{mn}, ϕ, ψ (the wave-function)

16.1. Kaluza's dimensional explanation

Kaluza assumed that the values of the fields in **5-D** are "almost constant", i.e. the derivative of all fields are small $\partial_4 g_{mn}$. This needs to explain the illusion of major changes in the values of the fields $g_{\mu\nu}$ or A_μ in **4-D**.

Kaluza's theory is similar to the truncation model presented in Section 15.8. Kaluza minimizes the influence of the fifth dimension on the **4-D** world by assuming **CYL**.

Another analogy can be useful here. Think of adding time as the fourth dimension of spacetime structure. In classical mechanics we assume small velocities, i.e. that trajectories are such that $\partial_0 x$ are small compared to the unity. In fact, for photons the previous derivative always equals one. This is not in general true for all particles and all motions. The same problem haunts Kaluza's theory: why should we assume that all fields, i.e. all components of the field g_{mn} are small? This seems ad-hoc an assumption. In fact Kaluza's theory was a highly idealized model. He needed to use two physics for massive particles and microscopic particles.

Notwithstanding these difficulties, Kaluza was able to derive the Lorentz force acting on a charged particle in **5-D**. Kaluza explained the trajectories of **4-D** particles as projections of the **5-D** geodesics. He also managed to infer the trajectories of charged particles from the gravitational field g_{mn}. The field ϕ was also a source of major troubles for Kaluza because it generated an immense force, much greater than the one usually observed.

I take unintended explanations the consequences of step 1 which were not directly envisaged by Kaluza or by Klein when they unified **EM** and **GR**. There are few unintended explanations in Kaluza's theory, except his speculation that the ϕ field would render the quantum statistics fluctuation. I discuss this in the context of Klein's dimensional explanation. In short, Kaluza's explanations are consequences of the unificatory process and are less powerful and relevant than Klein's explanations. I also deem that Kaluza's result lacks the physicality of Klein's explanations as discussed in the following sections.

16.2. The heuristics of Klein's compactification

The main scope of the rest of this chapter is to show that Klein's compactification **COMP** explains more than Kaluza's cylindricity **CYL**. There are some specific features of Klein's hypothesis that are relevant to explanation.

I want to argue against the argument that Klein's **COMP** condition is a sheer change of the coordinate system. Compactification can be spurious and misleading.[272] In **GR** there is a distinction between genuine singularities and mere artifacts of coordinate choice. Similarly, one has to ensure that they belong to the real physical space and not merely to its

[272] It can be said that spacetime singularities have a similar history, from abhorrence to unconditional acceptance.

representation. Indeed, in a polar coordinate system (R,θ,φ), it is always the case that $\varphi = \varphi + 2\pi n$, where *n* is a natural number that can count the number of windings. One may be tempted to say that we are living in a compactified polar dimension φ, but that is not the case. The polar system of coordinates has a symmetry which is not a proper one because through a suitable transformation to Cartesian coordinates *(x,y,z)* the cyclical symmetry is removed and the system is isomorphic with $\mathbb{R}^3$ which does not have a cylindrical symmetry. Indeed, some polar coordinates look like compactified spaces but they are only apparent: there are transformations of coordinate that removes such a compactification. Klein's **COMP** condition rules out this possibility: there are no transformations that remove the symmetry S(1). **COMP** is not a mere new coordinate system of *representation*, but the structure of the *real* fifth dimension. Same argument can be run for the singularities of the polar coordinates at $r \to 0$ of the coordinates of $\mathbb{R}^3$, which is not a singularity of the spacetime, but only of the representation of the space time.

The analogy with the hydrogen atom. Klein used the analogy between the behavior of a wavefunction on a closed orbit in a hydrogen atom and the behavior of the wavefunction in **5-D**. The analogy has a pure heuristic role, as he has been inspired by early quantum results on closed orbits. The mathematical structure in both cases is of a periodic function and hence the idea of a Fourier expansion. But, again, while the hydrogen atom can be represented in a coordinates in which $\varphi = \varphi + 2n\pi$, the atom itself does not live in a compactified space. Though it is not simply a classical "quantity of motion", quantum mechanical momentum has some properties of classical mechanical momentum (associated to moving particles or to waves). But the **QM** momentum sometimes has a discrete spectrum, i.e. it is quantized. As p^4 depends linearly on e, which is quantized, one may ask whether it is quantized, too. In polar coordinates, $\dot{\phi}$ or $\dot{\theta}$ are velocity-like quantities (they are actually angular velocities and there is an "angular momentum"), whereas p^4 is different. One can see the dynamics on x^4 as generating the illusion of charge quantization.

16.3. What does Klein's compactification of x^4 explain?

There are two classes of "dimensional explanations" that Klein has achieved as direct consequences of the unificatory procedure. First there are some *intended* explanations. Similar to Kaluza, his intention was to explain the **4-D** appearance of gravity and electromagnetism as a projection of a **5-D** like gravity. Fields in **5-D** dimensions are decomposed in lower-dimension fields by a procedure similar to a projection. Klein looked for the projection of the **5-D** geodesics onto the **4-D** space for various types of charged particles. These projections worked as expected, although many aspects of them are unclear even today. Indeed, in the presence of electromagnetic and gravitational fields and in the absence of other material fields, the real trajectories of probe particles coincide with the geodesics of the Klein metric in $\mathbb{M}^4 \otimes S_1$.

If trajectories were intended explanation of both Kaluza and Klein theories, some other facts were unexpectedly explained away. For Kaluza, as well as for **EM** or **GR**, the charge quantization, the symmetry of **EM** and the existence of some particles were brute facts, whereas in Klein's theory they become *explananda*. Once one has accepted **`COMP`**, one hits the ground of explanation and no explanation is needed anymore. **`COMP`** is a brute fact at this stage of the theory. The "unexplained explainer" is that the fifth dimension is curled and this is for Klein a brute fact such that no other *explanans* is necessary. I argue that that there are several advantages of Klein's theory as a unificatory theory besides it being parsimonius and economical: it explains, it can solve problems and more remotely it can offer would-be predictions. At a more general level, both Kaluza and Klein taught us how to reconceptualize the dimensionality of spacetime through the unificatory looking glass.

Klein's reversed argument, in which **COMP** becomes a brute fact that explains **CYL** provided Klein with a powerful unificatory mechanism able to generate some important *novel* explanations. His result surpassed his original expectation by explaining the quantization of the electrical charge and the internal symmetry of **EM** as the symmetry of *S(1)*. In addition, there were other unintended, albeit less successful, explanations in Klein's theory.

This projection has also some unexpected and unintended consequences. The most important is the quantization of charge and some indications about the mass of the photon. In a hindsight, by its interpretation of the dynamics on the fifth direction, Klein made room for the existence of positron, too: a particle with the same mass as the electron, the same charge, but positive. It is time now to investigate how Klein's dimensional explanation provides answers to why-questions: "why is the charge quantized?"; "why does the **4-D EM** have the symmetry U(1)?"; "why is the mass of the photon zero?" or question regarding the existence of not yet observed phenomena such as the positron or the graviton.[273]

[273] The positron was detected for the first time in 1930 by C.-Y. Chao and by C. D. Anderson in 1932 who gave them the current name. The existence of a graviton has not yet been confirmed, although some current theories predict its existence.

I want to discuss here the unintended explanations in Klein as they constitute the novelty.

16.4. The explanation of charge quantization: an unexpected *explanandum*

In classical **EM** there is an experimental fact:

[90] *The electrical charge is quantized*

which is brute in the sense that it is not explained by the classical **EM** theory. Even after the advent of Q**ED** stayed as a “brute” fact of physics. Despite the same word, quantization of charge and the quantization of the **EM** field in **QED** are two independent subject. You can have a classical theory with quantized charge or the other way around.[274]

In fact few theories tries to explain the quantization of charge: almost all physical theories take it a a brute fact, except Grand Unified Theory which claims that it can explain the quantization of charge.[275] Few theories except those assuming extra spatial dimensions can explain the quantization of charge and Klein was able to explain it as a consequence of **COMP** and the hypothesis of wave mechanics.

The second part of the 1926 paper and the note to *Nature* are directly connected with two major developments of both relativity and quantum mechanics: Schrödinger equation and de Broglie’s hypothesis of the pilot wave. In his memories, Klein reminisced:

---

[274] One might say that quarks have fractional charge. One way to dismiss this statement is to say that quarks in themselves do not contribute to the charge of macroscopic objects. The charge of body is computed as the charge of all electrons within it.

[275] The values of electric charge operator **Q** is $\mathbf{Q} = \mathbf{T_3} + \mathbf{Y}$. In **GUT,** both $\mathbf{T_3}$ and **Y** operators are embedded in a SU(5) simple group, hence the values of **Y**, like those of $\mathbf{T_3}$ are constrained by the structure of the algebra, hence the the quantization of the electrical charge has been derived. In the last decade **GUT** is believed to be dismissed as empirically adequate by experiments with the proton decay. See details in (Georgi 1999, 320)

> [I tried] to learn as much as possible from Schrödinger and also from de Broglie, whose beautiful group velocity consideration impressed me very much even if by and by I saw that it did not essentially differ from my own way by means of the Hamilton-Jacobi equation. From Schrödinger I learnt in the first place his definition of the non-relativistic expressions for the current-density vector, which it was then easy to generalize to that belonging to the general-relativistic wave equation. In this, after Schrödinger's success with the hydrogen atom, I definitely made up my mind to drop the possible non-linear terms, although I was still far from certain that this was more than a linear approximation. Also I derived the energy-momentum components, which in the five-dimensional formalism belonged to the current-density vector. These I published much later, due to the appearance in the meantime of a paper by Schrödinger containing the corresponding non-relativistic expressions (Ekspong 1991 111-112).

Klein explicitly relied on de Broglie's treatment of quantum phenomena by analogy with mechanics. Especially in his PhD thesis (1924), de Broglie's associated to each bit of energy with mass m_0 a periodic wave with a wavelength: $\nu_0 = \frac{m_0 c^2}{h}$. The group velocity of this wave is the same as the velocity of the mass (de Broglie 1924, 111, [1]). Previous physicists' experiences with de Broglie's model were successful: Sommerfeld's condition for stability on hydrogen orbit if one assumes the conservation of phase. Klein extended de Broglie's hypothesis to **5-D**: reality is described by a **5-D** "pilot" wave function. He started from the analysis of geodesics and hypothesized that according to de Broglie, geodesics should be rays of the wave function (Klein 1926 900-902). In other words, in **5-D** geodesics of elementary particles and the rays of the associated waves should be identical.

This facilitated the explanation of [90]. The extrinsic element of unification is the behavior of the wavefunction which provided Klein with a clear form of a momentum on the fifth axis. The "momentum" on x^4 and it can be interpreted as electrical charge q/m. In this sense, the momentum has a non-dynamical interpretation. Though it is not a "quantity of motion”, it has some properties of a momentum (always associated to moving particles or to waves). Because x^4 is compactified, it is natural to take p^4 as quantized, too. What is the momentum on this axis? Klein wrote in *Nature*:

> The charge q, so far as our knowledge goes, is always a multiple of the electronic charge e, so that we may write $p_4 = n\frac{e}{k}$ with $n \in \mathbb{Z}$. This formula suggests that the atomicity of electricity may be interpreted as a quantum theory law (Klein 1926, 516).

As the definition of momentum is given by:

$$p^4 = \int dx^4 T^0 = \int dx^4 g^{m0} g^{n4} (\partial_m \varphi)^* \partial_n \varphi \qquad (152)$$

one can assume that:

$$p^4 = const / \lambda_4 \qquad (153)$$

More precisely:

$$p_4 = ne / c\sqrt{2\kappa} = n\hbar / \lambda_4 \qquad (154)$$

where κ is a constant related to G, the constant of universal attraction, see (115). From this we can infer the quantization of the charged particle as being imposed by **COMP**. The particle cannot have an arbitrary value of the p^4 momentum. This means that if the fifth dimension is compactified with a period of $2\pi\lambda_4$, then the electrical charge appears *quantized* in **4-D**.

The elementary electrical charge was experimentally known since Millikan’s experiments in 1911. What was not known was the compactification factor. From the simple linear relation Klein calculated the compactificaton radius: $\lambda_4 = 0.8 \cdot 10^{-30} cm$. The smallness of λ_4, which is less than the Planck length, is the reason to why extensions on x^4

cannot be observed by macroscopic observers. One can see that this is one of the intended explanation of **COMP**, as previously intuited on completely different grounds by Hinton.

As a second consequences of **COMP**, Klein realizes that the discreteness of the charge spectrum, via the de Broglie relation, leads to a discrete wavelength in the fifth direction. The particle's momentum in the fifth direction is a rest mass in four dimensions, since it moves along a five dimensional null geodesic and thus it does not have a rest mass in **5-D**. The mass is of the order of the Planck mass. Moving along x^4 is not simply a mechanical change of coordinates. Klein interpreted the fifth axis by looking at the "initial conditions" and at the concept of geodesics. If two particles have the same initial condition in **4-D**, x_0^μ but different ratios *q/m*, they will fall under the same geometrical shape in **5-D** by following the same trajectories. In this case the wave equation in *is* the equation of a null geodesic in **5-D**, i.e. $P_m P^m = 0$.

A third consequence of Klein's theory is the physical possibility of the existence of positrons. If a particle has the same velocity u^4 as an electron but in the opposite direction, it manifests in **4-D** as a positron. Then positron is like the electron, but they had different initial conditions on the direction x^4. Positrons were predicted by Dirac on independent grounds in 1928 and detected in the early-1930s. Gravitons and photons are excitations of the g_{mn} field but projected on the **4-D** subspace. I take all these as unintended explanations and predictions of Klein's theory. Klein never mentioned positrons before Dirac, but it is fair to mention here that positrons, gravitons and photons are logical consequences of Klein's theory. Klein's argument can in principle predict the existence of other particles, but the resources are limited. To predict other particles with spin for example, some supplementary fields needs to be added to g_{mn}.

Are brute facts reduced in Klein? If we think that **COMP** is a feature of the **5-D** spacetime, one can see that Klein was able to explain geometrically the some facts of the electromagnetic world, most notable the quantization of charge by geometrical means. The new brute facts are now **COMP**, i.e. the compactified space S^1 and some hypotheses related to the behaviour of wavefunction in such spaces. A simple analysis would reveal that Klein does not reduce the number of brute facts. Klein hypothetizes that the fifth dimension is curled and from here he inferred some unexpected consequences.

This can generate some philosophical dissatisfactions. Klein replaces a testable and empirical observable fact about the quantization with a “brute fact” about an unobservable, the fifth dimension of spacetime (I doubt this can be named a fact anymore). This could look fallacious and can be deemed as philosophical dubious. C. Callender proposes a way to see what facts are brute or not: “What we do not want to do is posit substantive truths about the world a priori to meet some unmotivated explanatory demand—as Hegel did when he notoriously said there must be six planets in the solar system.” (Callender 2004 206). One can see that Klein’s hypothesis is able to explain the symmetry of **EM** and in some respect (and under important qualifications) the quantum program. The hypothesis of a compactified dimension is less *ad-hoc* than it seems and also can be taken as a brute fact at least in this stage of the theory. There is nothing wrong with assuming something a priori: but in general we might want to know whether what we assumed is a substantive truth.

What about Klein's assumptions? Taking seriously extra dimensions of spacetime was assumed in late-1900s. Klein showed how properties of particles can arise from the extra dimensions of spacetime. String theorists discovered that in six dimensions Klein's theory can explain much more than it did in five dimensions. They rolled the extra dimensions in "some manner" and use the interpretation of motion in new directions to explain "the internal machinery of elementary particles" (Susskind 2005 235). In String Theory particles are replaced with strings or branes. The difference is that unlike particles, string can do something that a particle cannot: they wind around the cylinder: but this assumption about strings adds an important parameter to the theory, non-existent in Klein: the "winding number".

16.5. Quantum mechanics and the Kaluza-Klein unification

Klein's original intention had been to unify **EM** and **GR** but his new assumption on the structure of the fifth dimension surpassed the original aim: his approach constituted a first step toward the unification of general relativity with the formalism of quantum mechanics (which at that time had still been *in nuce*). In this sense, he targeted finally the worst dualistic nightmare ever according to which we need two formalisms, one for matter (as described by **QM**) and one for fields (**GR** and **EM** or their unification or whatever theory we have). *In toto*, Klein's project failed. But his theory has the resources to provide definitions for three elementary particles. For him, as for de Broglie, material particles are solutions to fields and their motion reflects the propagation of waves: "the observed motion as a kind of projection onto space-time of a wave propagation taking place in a space of five dimensions." Klein showed how Schrödinger equation could be derived from the wave equation in **5-D** in which:

> In a former paper the writer [Oskar Klein] has shown that the differential equation underlying the new quantum mechanics of Schrödinger can be derived from a wave equation of a five-dimensional space, in which $\hbar$ does not appear originally, but is introduced in connection with the periodicity in x^4. Although incomplete, this result, together with the considerations given here, suggests that the origin of Planck's quantum may be sought just in this periodicity in the fifth dimension. (Klein 1981 [1926] n)

Does Planck's constant indeed originate in the periodicity of the fifth dimension? Unfortunately, this is only a partial result—at best. One can infer some quantum numbers, especially the quanta of charge, from the symmetries of x^4, but not all of them. How much of quantum theory can be explained by this geometrization program? Not much. Quantum theory in its Hilbert space formulation is not captured by the topology of the fifth dimension, so one should have serious doubts about whether the whole quantum theory can be derived from topological assumptions in extra dimensions.[276] In the eyes of today's physicist, Klein's deduction is flawed: the *classical* theory of fields, even in **5-D**, is not able to provide a description of quantum phenomena. The strong intuition is that **COMP** is simply not enough to explain the whole **QM** theory. It is worth noticing that Klein would need an independent derivation of the scale of the compactified dimension.[277] One can see that **COMP** relies on the quantization of charge and on the interpretation of the p^4 as a ratio of e.

A more general question is whether **QM** be derived from a geometrization program. Initially, Klein intended to use quantum formalism in geometrization to infer the quantization of charge. The perspective of providing a ground for the quantum program

[276]The question whether a 5-D theory can capture the description of other interpretations of quantum mechanics (Bohmian mechanics, for example) is way beyond the scope of the present dissertation.
[277] Thanks to Christian Wüthrich for making this clear.

stemming from the geometrization of the fields tantalized Klein: “The strong impressions this [unification] made on me came from the attempt to find a wave background to the quantization rules.” In an interview taken by Th. Kuhn in 1962 Klein said: “In earlier years Bohr himself—and that played a role for me—had said that since you cannot get a connected picture of quantum phenomena and four dimensions that maybe you could in a higher number of dimensions.” In another interview he declared:

> I remember I was thinking of the fifth dimension already in the summer [of 1922] in Göttingen... [In the autumn of 1924] I had the main idea of wave mechanics. It was only a sketch on a few sheets of paper, but I could not find it later on when I wanted to find it. It may have been left in Ann Arbor... Then I was trying to find the stationary states of the harmonic oscillator. But I knew too little about the mathematics there, so I had not found it when Schrödinger’s work came about the hydrogen atom. (Pais 2000 132)

He tried to reverse this dependence and to derive **QM** from geometrization such that the wave function in **4-D** can be inferred from the wave function in **5-D**. Such a derivation of **QM** from geometrization could have been a huge victory for the classical theory over the still-to-be-born **QM**. In the late 1920s Einstein had expressed on various occasions skepticism over **QM**. Quantization and singularities were not at all trinkets on Einstein’s favorite menu list.[278] But the unification programs based on “classical fields” were already in a decline in 1927. After a number of disappointments with following Weyl’s and Eddington’s theories of affine metrics, Einstein returned to Kaluza’s paper in 1927. This year also marked an essential change in Einstein’s thought. Instead of looking for non-singular solutions to his equations as elementary particles, he started to think of elementary particles as singular solutions of the field equations. Einstein still believed that

[278] It is not here the place to describe his uneven position against the quantum program and his strong belief that quantum physicists abandoned the principle of localization and the natural laws in their causal form.

the field description could explain the "discrete" aspects of matter, including the dynamical equations of motion of elementary particles usually derived through quantum mechanics. In his "Allgemeine Relativitätstheorie und Bewegungsgesetz" (1927) Einstein's field equations do indeed contain the law of motion of singularities provided that the nature of singularity is specified only as a first approximation. His ultimate hope back then was that a Kaluza-Klein type formalism would explain finally **QM**. Einstein's own relation with the **5-D** was tempestuous: around 1940 he was convinced that the fifth dimension was totally artificial. Einstein's final words on Kaluza-Klein formalism hint toward a regression to the previous form of relativity in which the electromagnetic field is *not* geometrized. In his later years, Einstein endorsed again the treatment of the **EM** field in **GR** as part of the "wooden" part $T_{\mu\nu}$ as long as the space is free of ponderable matter and electric charges. The Maxwell field equations in a covariant form and the standard **GR** can provide sufficient differential identities to guarantee their reciprocal consistency. The two theories can peacefully coexist as long as they describe only fields in a spacetime free of matter.

> The desire to have the greatest possible unity has resulted in several attempts to include the gravitational field and the electromagnetic field in one formal but homogenous picture. Here we must mention particularly the five-dimensional theory of Kaluza and Klein. Having considered this possibility very carefully I feel that it is more desirable to accept the lack of internal uniformity of the original theory, because I do not consider that the totality of hypothetical basis of the five-dimensional theory contains less of an arbitrary nature than does the original theory. (Einstein 1949 84)

Einstein tried to compare the assumptions of the Kaluza-Klein theory with the assumptions of **EM**+**GR** and concluded that they are not enough reasons to choose Kaluza-Klein. After his initial enthusiasm with Kaluza's theory, Einstein deliberately ignored the question: how does Klein's theory connect with **QM**? The answer is a deferring one. Klein showed that one could infer the quantum numbers, especially the quanta of charge from the symmetries of the extra dimensions. Then, there is the “photon”. Einstein wanted quantum theory and all quanta, photon included, to emerge from one of the geometrized theories. Einstein did not want to “postulate” quanta or to impose them upon the geometrized theory. If there are photons, one should expect sources of gravitational field. Geometry has to be quantized, and not the other way around—explaining the quanta from classical (Riemannian) geometry. The direction of reduction between geometry and quantum theory is still a puzzling issue which befuddled physicists for decades:

$$\text{geometrization} \rightleftarrows \text{quantization} \tag{155}$$

In **QFT** this ambiguity is broken as quanization has priority over geometrization because quantum fields only have a dual nature. One may ask how real is then the unification of **EM** and **GR** within a geometrization program if geometrical objects are dispensable. Einstein was looking for a mathematical object to represent electromagnetic field, other than the curvature. For Kaluza and Klein this other mathematical structure was the fifth dimension, while keeping the Riemannian curvature in **4-D**.

Nevertheless, we know that Klein had some internal resources to show why the fifth dimension was unobservable but seemingly, Einstein simply neglected this explanation. In the late ‘30s and ‘40s he devoted all his attention to developing a unified theory based on a non-Riemannian metric. What is strange and discouraging is that this theory was able to prove neither Maxwell’s equation, nor Einstein’s Field equations. In this case it can be said that **KK** scored better than any of the non-Riemannian formalisms on various issues: connection with quantum mechanics, explanation of the non-observability of the fifth dimension, unification of **EM-GR**.

Klein aimed to a unification in which **EM** symmetry is eliminated. Is this process fulfilled in the case of **EM** and **GR**? No: the geometrization of dynamics is incomplete. Even the most primitive Kaluza-Klein theory needs non-geometrical fields "coupled" with the metric. These "matter fields" seem ad-hoc for the point of view of the geometrization program. They are added to the "marble" in order to explain away the wood but they are of a wood nature as it were.[279] The other way to eliminate matter fields and to remove "the second order duality" is to relate these matter fields to supersymmetry, so this takes us back to geometry.

Kaluza-Klein theories sidesteps simplicity and indicates maybe that the geometrical reduction is not fundamental. And the matter fields actually ruin the beauty and the aim of the theory. A. Salam studied these excitations for a Kaluza-Klein type of theory and concluded that the Ansatz for the **4-D** manifold associated with the graviton, the Yang-Mills vector (or simply the photon for **EM**), and other types of scalars (Brans-Dicke for Yang-Mills or the "dilaton" for Kaluza-Klein) arise as "leading terms" in the expansions. The massive excitations have all spin 2 and they can be assigned to infinite dimensional representations of the non-compact group SO(1,2). These non-compact symmetries are spontaneously broken and are nothing more than spectrum generating terms (Earman 1972; Salam and Strathdee 1982).

---

[279] Other less successful attempts of Wheeler are known under the name of Geometrodynamics. See (Wheeler 1962, 334; Gruenbaum 1973, 775-800; Earman 1972, 634-647)

The final word on this issue is that one should have serious doubts on whether **QM** can be derived from topological assumptions about **4-D**+ dimensions. In the eyes of modern physicists the meaning of Klein's deduction is flawed: the *classical* theory of field, even the one in **5-D**, is not able to provide a description of quantum formalism. We know now that a some specific hidden variable quantum theories are not possible—Bell's inequalities as well as Kochen-Specker theorem or Gleason's theorem are recent results in this respect. It is a different question whether a quantum theory with extra-dimensions of spacetime is such a hidden variable theory or not and whether various incarcations of the above results apply or not.[280]

16.6. Particles as *explananda* in Klein

In Klein, **`COMP`**, a *geometrical* brute fact, explains and predicts *physical* facts. Klein's aim was higher when he envisaged explaining particles. However, can a vacuum theory predict the existence of particles? His theory produced another unexpected explanation: the photon and, albeit Klein was not aware of it, the graviton and the "dilaton" could be deduced from `COMP` as expectation values of $\langle A_\mu \rangle, \langle g_{\mu\nu} \rangle, \langle \phi \rangle$ granting some first-order approximation: all massive states are disregarded—similar to the "dimensional reduction" used in modern Kaluza-Klein theories with D=11 by Scherk, Julia and Cremmer in 1978.

For many the **5-D** wavefunction comes with its own troubles: a tower of massive, charged and spin particles with mode n>1 having the mass $\mathrm{m}_n = |\mathrm{n}|\mathrm{m}$ pops into existence.

[280] I do not plan to enter into the details of this interesting issue.

In its original formulation, Klein's theory was not renormalizable.[281] Klein's world with a curled x^4 is operationally indistinguishable from a **4-D** world with an infinite mass spectrum. The renormalization is possible if we assume an energy cutoff point such that all energies are much smaller than a given quantity proportional to the inverse of the compatification radius $E \ll \frac{1}{\lambda_4}$. In a strong reading, if we want to adopt a more realistic attitude towards extra spatial dimensions we need to renormalize the theory: roughly speaking we need to show that quantities over a specific limit are not phyisical. In a more modest reading the theory can stay non-renormalizable and we still can study and classify all divergences without being too realistic about the extra spatial dimensions. For other theories such as **QFT** it can be shown that the need for renormalization is not intrinsically inexplicable (Huggett and Weingard 1996, S159). Such a result can be inferred perhaps for a Kaluza-Klein theory, too (Álvarez and Faedo 2006). What I want to suggest is that not being renormalizable is not in itself the death of a physical theory and that Kaluza-Klein theory is not better off that lots of other theories in contemporary physics.[282]

The "dimension reduction" is necessary to avoid embarrassing predictions. But in order to explain *massive* particles, one needs non-geometrical fields "coupled" with the metric, which indicates that the geometrical reduction is not fundamental. Despite Klein's attempts, "matter fields" must remain on the brute facts side and cannot be explicated away.

---

[281] One can associate these massive multiplets with the symmetry group of the theory. According to Salam, the non-compact symmetries are spontaneously broken and they are nothing more than spectrum generating terms (Salam and Strathdee 1982, 316-352).

[282] Thanks to Chris Wüthrich for asking this question on several occasions.

When fields and topology of spacetime are the variables, the dimensional truncation is called "dimensional reduction". This procedure was used in modern Kaluza-Klein theories with D=11 by Scherk, Julia and Cremmer in 1978. The authors started from a D=11 theory on which they put the "simplest set of fields" (Cremmer, Julia, and Scherk 1981 204). The optimal dimensional reduction is when the **4-D** known facts are recovered: symmetries of our spacetime, equations in **4-D**, several fields as we know them. The context was similar to Kaluza-Klein theories. The authors used the conjunction of:

[91] ***COMP:*** *compactify some spatial dimensions (see Section 11.3)*

[92] *Disregard the massive states.*

These two conditions are logically independent. There is a possibility of keeping only [92] and ignoring [91]. In this case extra "large" dimensions arise as alternatives to compactification . This is a reason to keep the two conditions logically independent.

Fourier expansion. In Klein's days, fields such as $g_{\mu\nu}(x)$, $A_{\mu}(x)$ or $\varphi(x)$ were thought to be mathematical objects which transform under four-dimensional general coordinate transformations. What Klein did not notice in 1926 but was consequently used by Einstein and Bergmann in 1938 is that if **COMP**, then all fields are periodical on x^4 and consequently they can be Fourier expanded having all other **4-D** fields as coefficients. This means that there is a duality between the "real" **5-D** tensors, vectors or scalars and their **4-D** "representations". Any field (scalar, vector or tensor in **5-D**) can be represented or decomposed in a Fourier expansion depending on an infinite number of its **4-D** components ${}^{(n)}g_{\mu\nu}(x)$. In this respect, the values of a field on the fifth dimension are reducible to an infinite number of values on four dimensions. Moreover, the first term of expansion or the expectation values of the field is independent of the fifth coordinate.

In the case of Klein's theory, because of **COMP**, the Fourier expansions of all fields living in the **5-D** manifold are:

$$\begin{cases} g_{\mu\nu}(x,y) = \sum_{n=-\infty}^{n=\infty} {}^{(n)}g_{\mu\nu}(x)e^{iny/\lambda_4} = \\ {}^{(0)}g_{\mu\nu}(x) + {}^{(1)}g_{\mu\nu}(x)e^{iy/\lambda_4} + {}^{(2)}g_{\mu\nu}(x)e^{i2y/\lambda_4} + \dots \\ \\ A_{\mu}(x,y) = \sum_{n=-\infty}^{n=\infty} {}^{(n)}A_{\mu}(x)e^{iny/\lambda_4} = \\ {}^{(0)}A(x) + {}^{(1)}A_{\mu}(x)e^{iy/\lambda_4} + {}^{(2)}A_{\mu}(x)e^{i2y/\lambda_4} + \dots \\ \\ \varphi(x,y) = \sum_{n=-\infty}^{n=\infty} {}^{(n)}\varphi(x)e^{iny/\lambda_4} = \\ {}^{(0)}\varphi(x) + {}^{(1)}\varphi(x)e^{iy/\lambda_4} + {}^{(2)}\varphi(x)e^{i2y/\lambda_4} + \dots \end{cases} \qquad (156)$$

where x is all x^{μ} and y is only the x^4 (for notational purposes in this section $y{=}x^4$)

Gravitons, dilatons, photons. Now it is time to discuss [92] and its consequences. What are the *explananda* of the Fourier analysis on the S^1? Firstly, it is the mass of the photon. Although the interpretation of zero modes as masses was too bold for the 1920s, Klein inferred the mass of the photon. If charge is the component of momentum in the x^4 direction and if one associates a wave to the motion of such a charged particle, the mean value of the field A_{μ} can be associated to the mass of the photon and $\hat{A}^n$ becomes a creation operator after quantization. It is clearly an idealization purporting that the fields do not depend of the fifth dimension x_4 because only the first term (n=0) in the Fourier expansion counts. In some cases it can be a reasonable approximation, but in other cases, Klein's theory included it looks arbitrary. It says that we neglect too *massive* particles, which is counterintuitive.

Even if the interpretation of mode zero as masses was too bold for the 1920s, I added all these results as unintended *explananda* of Klein's theory. Much later, in the 1980s, the expectation values of these fields: $\langle g_{\mu\nu}\rangle, \langle A_{\mu}\rangle, \langle\varphi\rangle$ given by the first terms in the Fourier series have been interpreted as masses of particles. As Duff remarks, in today's particle parlance, the Fourier coefficient of order zero describes a graviton (spin 2), a photon (spin 1) and from (spin 0) the dilaton (Duff 1995 6)**.**

It is easy to see now that the Fourier expansion is explanatory: the masslessness of graviton $\langle g_{\mu\nu}\rangle = 0$ is due to the general covariance of **GR**, the masslessness of photon $\langle A_{\mu\nu}\rangle = 0$ is explained by the gauge invariance of **EM** and the masslessness of the dilaton $\langle\varphi\rangle = 0$ to it being a Goldstone boson. The dilaton refers to the arbitrary scalar field ϕ which obeys the non-linear equation:

$$\Box\phi = -\frac{\kappa^2\phi^3}{4}F^{\mu\nu}F_{\mu\nu} \tag{156}$$

This last conclusion was obviously not present in Klein as he assumed ϕ=1. It was recently discovered that in fact Klein's argument is flawed and that he should not be assumed in the action, but at the end of the calculations (O'Raifeartaigh and Straumann 2000).

The massive excitations. Even nowadays compactifying a S^7 torus as Scherk, Julia and Cremmer did is not a piece of cake and notorious difficulties arise. Their procedure of dimension reduction as well as other attempts to compactify extra dimensions is nevertheless based on the same principles as Klein's **COMP**. In a modern reading, Klein's stated in 1926 that given today's high energy experiments and the smallness of λ_4, the modes beyond n=0 are large enough to be inaccessible from our **4-D** world. The massive modes are excited only in **5-D** and do not have observable consequences in **4-D**. Only the fundamental mode is visible from **4-D** and consequently the field values in **5-D** do not depend on the fifth coordinate x^4. This incorporates and in fact explains Kaluza's **CYL** condition. The series of massive particles is divergent. If we take into account the other terms of the Fourier series, a tower of massive, charged and spin particles all quantized invades the **4-D** world. Kaluza-Klein formalism is formed by an infinite number of massive particles with mode n. This is a shortcoming of the theory used to sustain from a formal point of view the skeptical stance against the **5-D** realism. Klein's **5-D** world with one of its dimensions compactified on a circle is operationally indistinguishable from a four dimensional world with a very particular (albeit infinite) mass spectrum. The infinite mass spectrum is a counterintuitive component of the original formalism, as well as of its subsequent developments.

In order to understand this problem we need to move from classical Kaluza-Klein theory to its modern developments. The strangest consequence of Kaluza-Klein and of its generalized forms remains the appearance of massive excitations as the expansion of the fields in terms of normal modes. It seems that it is necessary to introduce non-geometrical "matter fields" to act as sources in this allegedly geometrical theory. The interpretation of matter as modes of fields remains problematic and I need to leave this controversial aspect of Kaluza and Klein theories aside here.

16.7. Symmetries of EM as *explananda* in Klein

As I discussed in the context of unification, Klein was able to recover the internal symmetry U(1) of the **EM** from external symmetries of the spacetime by compactification, reducing thus the types of symmetries of his unified theory to only external symmetries (see Section 14.2 and 14.4). This result constitutes unification, but it is an unintended explanation, too: it explains an internal symmetry as an external symmetry, i.e. of spacetime. It is as if one can convert an internal symmetry into an external one by trading in a dimension of spacetime. The result shows us that once **COMP** is assumed we measure and perceive the U(1) of electromagnetism as a consequence of the spacetime structure.

If one wants to put a covariant field theory in a space with more than four dimensions, then one aims for a geometrical interpretation of internal symmetries in terms of external symmetries, i.e. the reduction of internal symmetries to external ones. This "geometrical" dream of symmetries lies at the foundation of Klein's theory. In many respects, its unification power and its generality come from the reduction of types of symmetries of the two theories. The resulting theory has only external symmetries whereas the original theories (relativity, electromagnetism and also non-Abelian type theories have altogether external as well as internal symmetries). This reduction of types of symmetries is another feature of the Klein's type of unification and it is specific to theories with extra spatial dimensions.

The internal symmetry of **EM** is explained away as an external symmetry of S^1 on x^4. By this procedure Klein reduces the types of symmetries, and not, strictly speaking, the number of symmetries. A brute fact such as "the symmetry of **EM** is U(1)" is explained as a symmetry of S^1. The new brute fact is now the symmetry of this manifold and not the internal symmetry of **EM**.

Perhaps it can be inferred by a weak induction that "carrying this logic to its ultimate conclusion, one might be tempted to conclude that there is no such thing in nature as an internal symmetry, even apparent discrete internal symmetries like charge conjugation being just discrete spacetime transformations in the extra dimensions. One can only speculate on how the course of twentieth century physics might have changed if, in striving towards non-abelian gauge fields in 1939, […] Klein had applied his own ideas to a sphere instead of a circle" (Duff 1995 15-16).

What are the advantages of converting internal symmetries to external symmetries (of spacetime)? One straightforward answer is parsimony and economy of thought. If all interactions can be described this way then String Theory or a program like geometrodynamics explains away the illusion we have about an inherent duality in all our physical theories: spacetime and matter. If this program cannot be carried on till the end, then we need to keep such a duality. Klein showed that for **EM** the internal symmetry can be explained away: similar results are true for the class of non-Abelian gauge theories. If this is the case for other forces or other fields goes well beyond Klein's theory: most likely, a pure geometrization project would never be complete. Klein's geometrical explanation is partial and works for a specific class of gauge theories.

16.8. The stability of Klein's vacuum

There is also another result in Klein which is clearly an improvement upon Kaluza: the stability of vacuum solutions: in short, Klein's vacuum is stabler than Kaluza's because it has the right kind of symmetry demanded in five dimensions.

Symmetries of theories and how to break them. I start with a simpler case of symmetry. Symmetries belong to theories in the sense that they are studied as symmetries of the Lagrangian most commonly used by the theory itself. Some theories have a symmetric Lagrangian, more precisely a Lagrangian invariant to a class of transformation. The symmetries of the Lagrangian are in general called gauge symmetries. For different systems one can add asymmetric terms and break the symmetry of the Lagrangian; this is not the main problem here. The real problem is that the solutions of the equations of motion derived from this Lagrangian by a variational principle have fewer symmetries. The vacuum state of the solutions, i.e. the broken symmetry state, is not invariant under the same symmetries of the underlying Lagrangian. The word "spontaneous" is infelicitous here, but it suggests that this symmetry breaking is not due to asymmetric terms in the Lagrangian, but to the asymmetries of the vacuum state (i.e. the lower energy solution of the equations of motion). In the case of the Higgs mechanism, the conventional wisdom is that the gauge symmetries of the Lagrangian are hidden in the broken symmetry state. The symmetries related to a theory can be then manifest or hidden.

The idea of spontaneous symmetry breaking is old in theoretical physics but it is a highly debated topic.[283] The conventional wisdom can be questioned on several grounds. For example, Penrose suspects that we have no empirical reasons to believe that the elec-

[283] The Higgs mechanism was developed in the 1960s by P. Higgs, Englert and Brout etc.

troweak symmetry, i.e. the product $SU(2)\otimes U(1)$ occurred as a physical process in the early stages of the universe (Penrose 2005 sect. 28.3). There are several reactions to Penrose's argument and it seems that we do have some empirical evidence that the break of the electroweak symmetry really occurred as a physical process.

Let us focus here on a different type of argument. In recent years, philosophers have started to pay more attention to the Higgs mechanism in which supposedly a symmetry is broken. Many criticisms are directed against some realist claims of the Higgs mechanism. M. Morrison questioned the Higgs mechanism in itself based on an analogy with Maxwell's ether and is illustrative of a general skeptical attitude against the existence of such a mechanism. J. Earman is skeptical about the claim that real massive boson, such as the Higgs boson can be created by eating the Nambu-Goldstone bosons, deemed "descriptive fluff" be Earman. His solution is to move from the Lagrangian description to the Hamiltonian description and drop altogether the gauge symmetry (Earman 2004, 1227; Earman 2003; Lyre and Eynck 2003, 277-303; Lyre 2001, S371). In short, the symmetry breaking mechanism is still largely debated among philosophers because it is not clear whether it is real, i.e. an event that occurred at an early stage of the universe or it is simply a drawback of our theories (Kosso 2000; Smeenk 2006; Earman and Morrison in Brading and Castellani 2003).

The unexpected result of Goldstone (1961) was that breaking a global symmetry of a Lagrangian had always the consequence the existence of a massless boson with spin zero. But there are no such bosons in nature. The immediate solution was to add mass terms to the Lagrangian that would destroy its global symmetry. From here, one can deny the existence of the Goldstone boson. The price to be paid is still high. The theory cannot be renormalizable—that is in itself a trouble maker. The second route, the one which is highly contested by the philosophers, is to accept that Goldstone bosons are "eaten" by the Higgs mechanisms.[284]

The simplest case is to look at a theory described by a scalar field depending on spatial dimensions. The very textbook example discussed in the philosophical literature is the so-called "Mexican hat" potential. The Lagrangian that describes the dynamics of the system can be split up into kinetic and potential terms. If you think of the kinetic energy as being represented as $\frac{mv^2}{2}$ or better $\frac{p^2}{2m}$ and by associating the derivative $\partial^\mu \phi$ and the potential energy by a term V, then the Lagrangian of a scalar field is:

$$\mathcal{L} = \partial^\mu \phi \partial_\mu \phi - V(\phi) \tag{157}$$

For a simple "Mexican hat" potential:

$$V(\phi) = -\mu^2 |\phi|^2 + |\phi|^4 \tag{158}$$

If $\mu^2>0$, there is only one minimum where the field is $\phi^* \phi=0$. But for $\mu^2<0$ the vacuum has more than one minimum given by:

$$\phi(\theta) = \sqrt{\frac{-\mu^2}{2}} e^{i\theta} \tag{159}$$

where θ take any value between 0 and 2π. The expectation value of this vacuum is $\sqrt{\frac{-\mu^2}{2}}$

The problem is that depending on the dimensions of the problem one can have an infinite number of states in three dimensions, two solutions for a two dimensions or one solution if we truncate the potential to positive values. Too many dimensions, i.e. too many

---

[284] I do not want to give there the full, quantum description of the Higgs mechanism. Textbooks provide the necessary technical details. For an excellent philosophical approach, see (Smeenk 2006).

symmetries create the illusion of a boson called the Nambu-Goldstone boson related to the "angular" degrees of freedom, i.e. the symmetry rotation of the hat. This boson does not exist. The two-dimensional symmetry that is associated to the vacuum with two values only gives rise to the Higgs boson which is seemingly real. Even in a classical context, we have here a problem of counting the mathematical/geometrical solutions in the sense of attributing the "right" number of physical solutions out of a given mathematical solutions. The "descriptive fluff", as Earman called it, needs to be separated from the real symmetries (Earman 2004, 173-198). This is of course a reason to look with a skeptical eye on such stories in which matter is created out of symmetry.[285]

Let us try there to connect this to Klein. In my analysis is important to see the relation among symmetries and dimensions. The particle sitting on the top of the hat can fall in the valley and it will stay there forever. The question here is what the vacuum solution is, i.e. the state with the lower energy, of such a problem. For the purpose of the analysis of the symmetries, adding an extra parameter, in other circumstances a benign parameter, can create more troubles than one expects. The point of the Higgs mechanism is that there is a mismatch between the symmetries of the Lagrangian and the symmetries of the solutions, or if you like, the observed symmetries. Similar to the Higgs mechanism, too many symmetries can bring in too many ghost entities. In the Kaluza-Klein unification more complicated fields are analyzed and put together in a specific spacetime structure with its own symmetry. Kaluza-Klein brings in a new dimension and one can ask is it the real

[285] The rest of Smeenk's analysis would take us too deep in the area of quantum field theory and too far away from the purpose of the present thesis; I confine here to the classical context of the Kaluza-Klein theories.

symmetry or people were drawn in by the apparent similarity of the gauge transformation of **EM** and those of parts of the gravity in **5-D**.

Think now of a "Klein mechanism". We start from a manifold with five dimensions with a field g_{mn}. We know that the vacuum solution of the **4-D** gravitational field, the least perturbed gravitational field is $\mathcal{M}^4$ or a Minkowski spacetime. We would think that the vacuum solution of the **5-D** field would be the Minkowski **5-D**, i.e. $\mathbb{M}^5$. I take one of Klein's major achievements the fact that he intuited that that cannot be the case. The original form of the action (97) would suggest the $\mathbb{M}^5$ vacuum (Aitchison 1991 162). But the factorized action used by Klein (Section 14.1) shows that the vacuum is not $\mathbb{M}^4 \otimes S_1$, which has less symmetries than the action we started with. If Klein lost or broke the symmetry, one can see here that unification is in fact made possible by this move from a symmetrical spacetime to a non-symmetrical one. Einstein disliked the lack of symmetry and he expressed it in an intuitive manner. Why is the fifth dimension special? Special here means different from the other three spatial dimensions.

There is a crucial point: in field theories the definition of energy depends on boundary conditions not only on the dynamics of the field. The previous analysis can be only partially useful. The boundary conditions of a Kaluza-Klein theory can in fact majorly change the story about symmetries. Focusing too much on the dynamics and forgetting the boundary condition can alter the analysis of symmetry. This is why it has been suggested that for spacetime theories we need to add a requirement (Blagojevic 2002 302):

> ***[93]*** *We can compare energetically only those solutions that have the same boundary conditions.*

The technical result not to be discussed further here is that only $\mathbb{M}^4$ has zero energy, so according to the definition of stability, only Minkowski spacetime $\mathbb{M}^4$ is stable.

Since the boundary conditions for $\mathbb{M}^4$ and $\mathbb{M}^4 \otimes S_1$ are not the same in Klein (again, in (Blagojevic 2002, 522) there is a clear confusion between Kaluza and Klein), then we cannot compare apple and oranges. From here, Witten showed that if the only criterion for the choice of the gravitational ground state is “stability”, then $\mathbb{M}^4 \otimes S_1$ is stable classically and semiclassically (Witten 1981b, 412). This result shows again why Klein’s solution is superior to Kaluza’s.

16.9. Valediction

I argued that in Klein’s case explanation is related to unification. I differentiate intended explanations from unintended explanations as well as direct and immediate consequences of Klein’s theory from indirect consequences. Unlike the models used in the previous chapter (see 15.8 and 15.9) I showed that explanation and unification are strongly intertwined in Klein.

On one hand, my argument is a counterexample to Morrison “decoupling claim” [12] (p. 124). I did not argue that all explanations in Klein are related to unification: I made a distinction between intended explanations which were part of the process of unification (through the process of identification) and unintended explanations which are consequences of unification but they were not part of the unification process in itself.

On the other hand, I argued here that Klein was able to explain phenomena well beyond the set of consequences of his unificatory assumptions. Take the explanation of the quantization of charge: it is based on **COMP** but it also based on the behavior of the matter wave in 5-D, i.e on its stationarity. This is a physical fact which was based on the analogy with the hydrogen atom. Contra Morrison, again, Klein's theory achieved explanation beyond its more narrow unificatory scope. Reducing the types of brute facts and helping us in understanding the role dimensionality of spacetime plays are two explanatory achivements ofKlein's theory (and in some respects, Kaluza's too).

To address again [56] (p. 263): I think there are interesting physical aspects of Klein's explanations that makes them novel and unexpected as well as genuine explanations. Physicality plays an important role in separating mere consequences of unification from genuine explanations: physicality or having a definite meaning is related to predictions. For example field ϕ does not qualify (yet?) as physical, compared to Klein's explanation of the quantization of charge. If Klein indeed predicted the existence of the positron and some features of photons and gravitons, then one can see that Klein's theory can predict some facts (the mass of the graviton or its existence are still highly depatable facts). Klein's theory offers some contrast examples of consequences versus explanations. The interesing and challeging task is to provide a more general answer to [56]: I claim here without entering into details that both physicality and novel prediction are part of separating mere mathematical consequences from genuine explanations.

I do not deny that there are mathematical explanations, but the geometrical explanation present in Klein have physical aspects which are not consequences of the mathematical formalism: the quantization of charge, existence of some particles and the internal symmetry of **EM** as consequence of the **COMP**. Some other results can be deemed as mere consequences of the fornalism: I reckon here especially the "explanation" of the trajectories of charged particles as projections of **5-D** geodesics. In conclusion, Klein's theory illustrates the coupling between explanation and unification and also the way in which explanations can go well beyond mere results or consequences of a unificatory formalism.

I surmise that this dissertation may generate two attitudes. It may look disappointing to somebody who is demanding immovable answers to the question about the reality of extra-spatial dimensions. In fact I talk about two theories which are most likely false. But philosophers who are intrigued by this hypothesis may find Kaluza and Klein theories rich and mature enough for a philosophical appraisal. They open the stage for philosophical interpretations and foundational studies. My proposal here was to look at these two theories through the glasses of scientific unification. I acknowledge that there are many other perspectives to look at them: as precursors of gauge theories, as precursors of the fiber bundle formalism etc.

With this dissertation, I fill a gap in the philosophical literature and incite further studies into their more elaborate incarnations.

Appendix. Differential geometry: metric and manifold

The manifold

A n-dimensional (n-D) spacetime manifold $\mathcal{M}^{(n)}$ is basically a set of objects that "looks like" open sets of $\mathbb{R}^n$, the set of n-tuples of real numbers.[286] This means that any point P has a neighborhood (an open set O^P of points that are close to P), that the spacetime is coordinatizable by $\mathbb{R}^n$ (we need only quadruples of real numbers to describe the position of a spacetime point, i.e. the spacetime can be charted by a map Φ from open sets O^P in the spacetime to $\mathbb{R}^n$) and that any for any function f in spacetime, its mapping onto $\mathbb{R}^n$, $f \circ \Phi^{-1}$ defined around the neighbor of a point P is differentiable for any chart Φ.[287] In order to define a topology on the $\mathcal{M}^{(n)}$, a relaxed condition is imposed upon all charts: they need to be homeomorphisms.[288]

All these requirements are local and remain local for the **GR**. For example the two dimensional sphere S^2 is a two-dimensional manifold only locally because it cannot be mapped globally to $\mathbb{R}^3$. Small neighborhoods around a point P can be mapped into $\mathbb{R}^2$, although this cannot be done for the whole S^2.

The metric as a tensor

In order to understand Kaluza's and moreover Klein's approach, I need to enter into some details concerned the metric. Intuitively the metric tell us something about how objects are related one to the other in the manifold. The line element ds^2 is associated to the

[286] Throughout my thesis an exponent in brackets refers to the dimensionality of the manifold.
[287] Charts are called "coordinate systems" in physics. It is evident that they are not unique.
[288] I use the terminology from (Wald 1984 13-28, 423-427).

way in which we measure space and time. Setting some mathematical details aside, the expression of the ds^2 is not rigorous. In fact, a more rigorous definition would leave the concept of line element aside and will emphasize the tensorial nature of *g*. The line element is still very useful and its expression can be related to the metric, but strictly speaking they are not one and the same.

Dual space

A dual space $\mathbf{V}^*$ is associated to any vector space and consists of all linear functionals $f: V \to \mathbb{R}$. The dual space of $\mathbf{V}^*$ is the space of all possible maps from V to $\mathbb{R}$. For example, if V is the space of row vectors $\begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix}$, V^* is the space of column vectors $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$, because their product is always a real number.

Tangent space

Let us think of a manifold $\mathcal{M}$ of dimension m embedded in $\mathbb{R}^m$. As $\mathcal{M}$ is differentiable, there is a family of curves passing through each point P of $\mathcal{M}$ and each of them carries its own tangent. The tangent space at point P is a vector space of dimension m and it is composed of all vectors that are tangent to all curves in $\mathcal{M}$ passing through P. There are some natural questions that arise of the tangent spaces:

- What happens when we change the coordinate system of the base manifold $\mathcal{M}$?
- What happens when we move from P to a point P' in its neighborhood? Are the tangent spaces related?

In order to understand these questions one need to introduce the concept of tensor, tangent space and dual space. In any vector space V one can define the directional derivatives operator associated to a vector *v*: $\sum_{\mu} v^{\mu}\left(\frac{\partial}{\partial x^{\mu}}\right)$ to a point P and a coordinate system $\{x^1 \ldots x^n\}$. From this one can build the tangent space V_p with coordinate basis $\left\{\frac{\partial}{\partial x^1} \cdots \frac{\partial}{\partial x^n}\right\}$. From the tangent space V_P one can construct the *dual* of the tangent space V_P^*. Its basis is $dx^1 \ldots dx^n$, where dx^μ is merely the symbol for the linear map from V_P to real numbers.

Building tensors

In general, a tensor of rank (k,l) is a multilinear map from k covectors in a dual vector space V^* and l ordinary vectors in V to real number:

$$T : \underbrace{V^* \times V^* \ldots V^*}_{k} \times \underbrace{V \times V \ldots V}_{l} \rightarrow \mathbb{R} \tag{160}$$

For example, a (0,2) rank tensor pairs two vectors and the result is a real number. A (0,2) rank tensor pairs two covectors. In general tensors of rank (k,l) need both the dual V^* and the ordinary V spaces. If $\{x_\mu\}$ is a basis of V and $\{x^{\mu *}\}$ is a basis of the dual space V^*, then a tensor can be filled in $n^{(k+l)}$ ways. The space of all tensors of rank (k,l) is a vector space having the dimension $n^{(k+l)}$. The numbers or the components of a tensor clearly depend on the two bases. Cases of tensors of low shows us that: a (0,1) tensor is a dual vector, a (1,0) tensor is a vector.[289]

[289] In fact is a double dual vector, but for all purposes the space V^{**} can be identified with V. See (Wald 1984 20).

One way to construct tensors is by taking outer products of vectors and dual vectors. These tensors are called simple tensors. Any covector can be decomposed then as: $\mathbf{V} = V_i \mathbf{d}x^i$ and any (0,2) cotensor is written as $\mathbf{F} = F_{ij}\mathbf{d}x^i \otimes \mathbf{d}x^j$. Any tensor is a sum of simple tensors. More generally, given a basis $\{v_\mu\}$ of a vector space V and the basis $\{v^{\nu *}\}$ of its dual space V^*, any (k,l) tensor can be expressed as a sum of simple tensors:

$$T = \sum_{\mu_1 \ldots, \nu_1 = 1} T^{\mu_1 \ldots \mu_k}{}_{\nu_1 \ldots \nu_k} v_{\mu_1} \otimes \ldots \otimes v^{\nu_k} \tag{161}$$

The metric

One needs the tangent space and its dual (called cotangent space) in order to define fields of tensors.

What is the metric *g*? The simple answer is: a tensor of rank (0,2) having the same dimensionality as the manifold and living in the tangent space V_P. The properties of the metric are:

- It is symmetric: $g(v_1,v_2)=g(v_2,v_1)$
- It is non-degenerate: For all v in V_P, if $g(v,v_1)=0$ then $v_1=0$.

How do we construct the metric? In today's parlance, we prefer to write the metric as a outer product of basic tensors $\boldsymbol{d}x^m$. More precisely, the metric is:

$$\mathbf{g} = g_{mn}\mathbf{d}x^m \otimes \mathbf{d}x^n \tag{162}$$

Where the $\mathbf{d}x^0 \ldots \mathbf{d}x^n$ is the canonical frame basis of vectors.[290]

The metric can be applied to dual vectors v^m and the result is a dual $g_{mn}v^n$. The inverse of the metric is usually a (2,0) tensor and for simplicity it is written g^{mn}. It applies to

[290] The outer product is omitted in standard textbooks. Its importance in Klein's metric will be discussed later in this chapter. For an interesting discussion about this omission, see (Wald 1984 23; Felsager 1998 322).

dual vectors $g^{mn}\omega_n$ and the result is a vector ω^n. The metric can be applied pair of vectors (v^m,w^n) and the result is a number: $g_{mn}v^n w^m$. As the metric tensor is always symmetric, the order in the pair does not matter.

For any metric and any point P there is always a orthogonal basis such that $g(v_\mu,v_\nu)=0$ and $g(v_\mu,v_\mu)=\pm1$. The pseudo-Riemannian metrics can have positive and negative numbers, but not zeros. The number of occurences of + signs and – signs does not depend on the choice of the basis and it is put in a pair (p,q). This is the signature of the metric and it gives the sign number of eigenvalues of the metric. Lorentzian metrics are always such that one value is negative and all the others are positive: (-1, n-1). A non-degenerate metric has no zeros in it.

In standard **GR**, spacetime is taken to be a 4-D differentiable manifold $\mathcal{M}^{(4)}$ and the metric to be pseudo-Riemannian.[291]

We know that **GR** is a theory formulated originally in a Riemannian space. Some of the Riemannian structures were endorsed with physical reality, some were considered mere representations. This is not a consensus in this respect. There is also a less evident link between **GR** and affine geometry, even if some could proclaim that in **GR** affine connections are use restrictively i.e. only local. The applications of generalized geometries to GR were first discussed by R. König, H. Weyl and Schouten, all in the early 1920s.

Even in a Minkowski spacetime, an inertial observer corresponds to a boost transformation in 4D. In this case the connections are affine and special relativity can be described by affine connections. In the case of curved spacetime, affine connections do not act globally. They can be applied to small neighbors only.

[291] Both Kaluza and Klein employed pseudo-Riemannian metrics.

Riemannian spaces

So at the very foundation of relativity lays the pseudo-Riemannian geometry which is already a non-Euclidean geometry. A pseudo-Riemmanian manifold is a smooth, i.e. differentiable manifold *M* equipped with a smooth, symmetric (0,2) tensor $g_{\mu\nu}$ which is not degenerate, but it can be negative. To any point P of $\mathcal{M}^{(n)}$ one can associate a tangent space. The tangent space is a *vector space* which intuitively contains the possible "directions" of all possible curves passing through P.[292] For example, the tangent space of a point P on a S^2 sphere embedded in $\mathbf{R}^3$ is simply the plane tangent to the sphere at point P. Indeed, the tangent space contains basically all possible derivatives of all possible curves on the sphere.[293]

In **GR** it is important to know how to "connect" two tangent spaces on curves in R^4 as it gives information on the dynamical evolution of system at two different times. The geometrical answer to this question depends on the metric of the original manifold. If the original manifold has no metric, then there are an infinite number of connections on that manifold, so practically there no dynamics to describe in such a space. If the metric is pseudo-Riemannian, the respective connection is called the Levi-Civita connection ∇. It is unique and it has some notable properties. For any three vector fields X, Y and Z:

- It is torsion-free , i.e. its torsion is zero:

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y] = 0$$

---

[292] The elements of the tangent space are called tangent vectors at *P*.

[293] As in general relativity the objects are often given without being embedded in a manifold, the tangent vector is defined as a directional derivative.

where $\nabla_X Y = X^\mu \left(Y^\nu \Gamma^\sigma_{\mu\nu} + \partial_\mu Y^\sigma\right)$ is the covariant derivative of vector Y on the X direction, given the connection Γ and [X,Y] is the Lie bracket of the vector fields X and Y.

- It preserves the metric g: $\partial_X (g(Y,Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$, i.e. has a parallel transport (it is an isometry).

How are two tangent spaces connected in a small neighborhood? The intuition would be that the connection between the two affine spaces should be a smooth one, i.e. an affine one. The theory of tensors on which **GR** depends is an affine theory. In any point x_0, one can define an infinite number of coordinate systems. In physics we are interested in those transformations $x \rightarrow \overline{x}$ related by smooth, maybe linear relations:

$$\overline{x}^\mu - \overline{x}^\mu_0 = \left(\frac{\partial \overline{x}^\mu}{\partial x^\nu}\right)_0 (x^\nu - x^\nu_0) \tag{163}$$

where μ=1...n.[294] The same can be said about the differential in the tangent space:

$$d\overline{x} = \sigma^i_j dx \tag{164}$$

which is again an affine transformation. The tangent spaces are always affine because the transformation of a vector V to a vector V':

$$V^{\mu'} = \Lambda^{\mu'}_\nu V^\nu \tag{165}$$

is affine, too.

In some specific sense **GR** endowed with the pseudo-Riemannian metric geometrized the gravitational field. Unfortunately, the geometrization of electromagnetism within a pseudo-Riemannian geometry is impossible because a 4-D Riemannian metric

[294] Where n is the dimension of the space, typically n=4 for **GR.**

cannot accommodate any other field than gravity. Put it crudely, mathematicians and physicists alike tried to show that general relativity employs the wrong geometry, or at least that the choice of geometry is arbitrary, at best. Others tried to show that the geometry employed in general relativity is not general enough. A logical consequence would be that a more general geometry could help to the unification of gravitation and electromagnetism. This is why one of the main reasons to move away from the Riemannian geometry was unification with other forces. Another logical consequence would be that an extra dimension is not necessary in order to achieve unification. Geometry can provide this in four dimensions only if the right geometry is employed.

Transformation of metric from one system to the other

If we change the coordinate system, the metric will change accordingly. As a (0,2) tensor, the metric obeys the general transformation rule. For a general coordinate transformation $\overline{x}^{\mu} = \overline{x}^{\mu}(x^{\nu})$, the change of the metric is:

$$\overline{g}_{\mu\nu} = g_{\rho\sigma} \frac{\partial x^{\rho}}{\partial \overline{x}^{\mu}} \frac{\partial x^{\sigma}}{\partial \overline{x}^{\nu}} \tag{166}$$

The Curvature

It is useful mentioning that curvature and covariant derivative are defined as tensors. The popular definition **(33)** suggests that a vector becomes a tensor. Indeed, in general, a derivative operator ∇T is a map that takes a (k,l) tensor $T^{m_1 \dots m_k}{}_{n_1 \dots n_l}$ and transform it into a (k,l+1) tensor. The result is denoted by $\nabla_c T^{m_1 \dots m_k}{}_{n_1 \dots n_l}$.[295] Given the derivative operator, one can integrate the notion of parallel transport of vector or tensors altogether in

[295] For the list of properties of the ∇ see (Wald 1984 31).

a very simple way. A tensor of arbitrary rank (k,l) is parallel transported along a curve C with tangent t^a such that the equation:

$$t^a \nabla_a T^{m_1 \ldots m_k}{}_{n_1 \ldots n_l} = 0$$

is satisfied along the curve. A vector V is transported from a point P to a point P' along a given curve C if at P' we have:

$$t^a \nabla_a v^b = 0$$

References

Aitchison, I. J. R. "The Vacuum and Unification." In *The Philosophy of Vacuum*, edited by Simon Saunders. New York: Clarendon Press, 1991.

Álvarez, Enrique and Antón F. Faedo. "Renormalized Kaluza-Klein Theories." (3 May 2006, 2006), http://arxiv.org/abs/hep-th/0602150 (accessed 8/20/2009).

Anderson, E. *The Campbell-Magaard Theorem is Inadequate and Inappropriate as a Protective Theorem for Relativistic Field Equation*2004, http://arxiv.org/abs/gr-qc/0409122v2 .

Appelquist, Thomas, Alan Chodos, and Peter G. O. Freund. *Modern Kaluza-Klein Theories*. Frontiers in Physics; 65. Menlo Park: Addison-Wesley Pub. Co, 1987.

Bailin, D. and A. Love. "Kaluza-Klein Theories." *Reports on Progress in Physics* 50, no. 9 (1987): 1087-1170.

Baker, Alan. "Quantitative Parsimony and Explanatory Power." *The British Journal for the Philosophy of Science* 54, no. 2 (June 1, 2003): 245-259.

Balashov, Yuri and Michel Janssen. "Presentism and Relativity." *British Journal for the Philosophy of Science* 54, no. 2 (June, 2003): 327-346.

Banchoff, Thomas. *Beyond the Third Dimension: Geometry, Computer Graphics, and Higher Dimensions*. Scientific American Library Series. Vol. 33. New York: Scientific American Library (W.H. Freeman), 1990.

Barnes, Eric. "Explanatory Unification and the Problem of Asymmetry." *Philosophy of Science* 59, no. 4 (December, 1992): 558-571.

Bartelborth, Thomas. "Explanatory Unification." *Synthese* 130, no. 1 (2002): 91-107.

Batterman, Robert W. "Critical Phenomena and Breaking Drops: Infinite Idealizations in Physics." *Studies in History and Philosophy of Modern Physics* 36, no. 2 (June, 2005): 225-244.

———. "Falling Cats, Parallel Parking, and Polarized Light." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 34, no. 4 (12, 2003): 527-557.

———. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford Studies in Philosophy of Science. Oxford: Oxford University Press, 2002.

Bechtel, William and Adele Abrahamsen. "Explanation: A Mechanist Alternative." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36, no. 2 (6, 2005): 421-441.

Bechtel, William and Andrew Hamilton. "Reduction, Integration, and the Unity of Science: Natural, Behavioural, and Social Sciences and the Humanities." In *General Philosophy of Science Focal Issues*, edited by Theo A. F. Kuipers. Amsterdam: North Holland, 2007.

Belot, Gordon. "Dust, Time and Symmetry." *The British Journal for the Philosophy of Science* 56, no. 2 (June 1, 2005): 255-291.

———. "Symmetry and Gauge Freedom." *Studies in History and Philosophy of Modern Physics* 34B, no. 2 (June, 2003): 189-225.

———. "Chaos and Fundamentalism." *Philosophy of Science* 67, no. 3 Supplement (2000): S454-S465.

———. "Why General Relativity does Need an Interpretation." *Philosophy of Science* 63, no. Supplement (1996): S80-S88.

Bergmann, Peter Gabriel. *Introduction to the Theory of Relativity*. Prentice-Hall Physics Series. Englewood Cliffs, N. J.: Prentice-Hall, 1942.

Bickle, John. "Revisionary Physicalism." *Biology and Philosophy* 7, no. 4 (October, 1992): 411-430.

Blagojevic, Milutin. *Gravitation and Gauge Symmetries*. Bristol: Institute of Physics Publications, 2002.

Boi, Luciano. "Theories of Space-Time in Modern Physics." *Synthese* 139, no. 3 (2004): 429-489.

Brading, K. A. and T. A. Ryckman. "Hilbert's 'Foundations of Physics': Gravitation and Electromagnetism within the Axiomatic Method." *Studies in History and Philosophy of Modern Physics.* 39, no. 1 (2008): 102.

Brading, Katherine and Elena Castellani. *Symmetries in Physics: Philosophical Reflections*. Cambridge, U.K.; New York: Cambridge University Press, 2003.

Brown, Harvey R. *Physical Relativity: Space-Time Structure from a Dynamical Perspective*. Oxford; New York: Clarendon Press, 2005.

Brown, Harvey R. and Oliver Pooley. "The Origin of the Spacetime Metric." In *Physics Meets Philosophy at the Planck Scale: Contemporary Theories in Quantum Gravity*, edited by Craig Callender and Nick Huggett. Cambridge: Cambridge Univ Pr, 2001.

Bub, J. "Quantum Mechanics as a Principle Theory." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 31, no. 1 (2000): 75-94.

Bunge, Mario Augusto. *The Myth of Simplicity; Problems of Scientific Philosophy*. Englewood Cliffs, N.J.: Prentice-Hall, 1963.

Burger, Dionys. *Sphereland: A Fantasy about Curved Spaces and an Expanding Universe*. Translated by J. Rheinboldt Cornelie. New York: Barnes&Noble, 1983; 1965.

Butterfield, Jeremy. "The Hole Truth." *British Journal for the Philosophy of Science* (March, 1989): 1-28.

Callender, Craig. *What Makes Time Special?* 2008, http://www.fqxi.org/community/forum/topic/302 .

Callender, Craig and Jonathan Cohen. "A Better Best System Account of Lawhood." *Philosophical Studies* (2009 forthcoming).

Callender, Craig. "Answers in Search of a Question: 'Proofs' of the Tri-Dimensionality of Space." *Studies in History and Philosophy of Modern Physics* 36, no. 1 (March, 2005): 113-136.

———. "Measures, Explanations and the Past: Should `Special' Initial Conditions be Explained?" *British Journal for the Philosophy of Science* 55, (2004): 195-217.

Callender, Craig and Jonathan Cohen. "There is no Special Problem about Scientific Representation." *Theoria: Revista De Teoria, Historia y Fundamentos De La Ciencia* 21, no. 55 (January, 2006): 67-85.

Campbell, John Edward and E. B. Elliott. *A Course of Differential Geometry*. Oxford: Clarendon Press, 1926.

Campbell, Joseph Keim, Michael O'Rourke, and Harry Silverstein, eds. *Causation and Explanation*. Cambridge, Mass.: MIT Press, 2007.

Cao, T. Y. and S. S. Schweber. "The Conceptual Foundations and the Philosophical Aspects of Renormalization Theory." *Synthese* 97, no. 1 (1993): 33-108.

Capri, Anton Z. and Abdul N. Kamal. *Particles and Fields, 2*. New York: Plenum Press, 1983.

Carroll, Sean. *Spacetime and Geometry: An Introduction to General Relativity*. San Francisco: Addison Wesley, 2004.

Cartan, E. "Sur les Variétés à Connexion Affine, et la Théorie de la Relativité Généralisée (Deuxième Partie)." *Annales Scientifiques de l'École Normale Supérieure Sér.3* 42, (1925): 17-88.

———. "Sur les Variétés à Connexion Projective." *Bulletin De La Société Mathématique De France* 52, (1924): 205–241.

———. "Sur les Variétés à Connexion Affine et la Théorie de la Relativité Généralisée (Première Partie)." *Annales Scientifiques de l'École Normale Supérieure* 40, no. 3 (1923): 325-412.

———. "Les Groupes Projectifs qui ne Laissent Invariante aucune Multiplicite Plane." *Bulletin De La Société Mathématique De France* 41, no. 1 (1913): 53-96.

Cartwright, N. "The Limits of Exact Science, from Economics to Physics." *Perspectives on Science* 7, no. 3 (1999a): 318-336.

Cartwright, Nancy. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press, 1999b.

Castellani, E. "Reductionism, Emergence, and Effective Field Theories." *Studies in History and Philosophy of Modern Physics* 33, no. 2 (2002): 251-267.

Chomsky, Noam. *New Horizons in the Study of Language and Mind*. Cambridge, U.K.: Cambridge University Press, 2000.

Churchland, Patricia Smith. *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Computational Models of Cognition and Perception. Cambridge, Mass.: MIT Press, 1986.

Churchland, Paul M. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, MA: MIT Press, 1989.

———. *Scientific Realism and the Plasticity of Mind*. Cambridge; New York: Cambridge University Press, 1979.

Clifton, R., J. Bub, and H. Halvorson. "Characterizing Quantum Theory in Terms of Information-Theoretic Constraints." *Foundations of Physics* 33, no. 11 (2003): 1561-1591.

Cohen, I. Bernard and George E. Smith. *The Cambridge Companion to Newton*. Cambridge, UK ; New York, NY: Cambridge University Press, 2002.

Corry, Leo, Jurgen Renn, and John Stachel. "Belated Decision in the Hilbert-Einstein Priority Dispute." *Science* 278, no. 5341 (November 14, 1997): 1270-1273.

Craig, William Lane. *The Tensed Theory of Time: A Critical Examination*. Synthese Library. Vol. 293. Dordrecht ; Boston, Mass.: Kluwer Academic, 2000.

Cremmer, E., B. Julia, and J. Scherk. "Supergravity Theory in Eleven-Dimensions." *Physics Letters B* 76, (1981): 409.

Darden, Lindley and Nancy Maull. "Interfield Theories." *Philosophy of Science* 44, (March, 1977): 43-64.

Darwin, Charles. *The Variation of Animals and Plants Under Domestication*. London: J. Murray, 1868.

David, Lindley. *The End of Physics: The Myth of a Unified Theory*. New York: Basic Books, 1993.

Davies, P. C. W. and J. R. Brown, eds. *Superstrings: A Theory of Everything?*. Cambridge U.K.; New York: Cambridge University Press, 1988.

Dawid, R. "Underdetermination and Theory Succession from the Perspective of String Theory." *Philosophy of Science* 73, no. 3 (2006): 298.

de Broglie, Louis. "Thèses Présentées à la Faculté des Sciences de l'Université de Paris: pour obtenir le Grade de Docteur Ès Sciences Physiques: Soutenues le Novembre 1924 Devant la Commission d'Examen." Paris, Masson, 1924.

Deser, S. "Dimensionally Challenged Gravities." In *Revisiting the Foundations of Relativistic Physics: Festschrift in Honor of John Stachel*, edited by John J. Stachel, Jürgen Renn, Lindy Divarci, Petra Schröter and Abhay Ashtekar. Vol. 234, 649. Dordrecht; Boston: Kluwer Academic Publishers, 2003.

DeWitt, Bryce S. *Dynamical Theory of Groups and Fields*. New York: Gordon and Breach, 1965.

Dickson, Michael. "Plurality and Complementarity in Quantum Dynamics." In *Scientific Pluralism*, edited by Stephen H. Kellert, Helen E. Longino and C. Kenneth Waters. Minneapolis, MN: University of Minnesota Press, 2006.

Disalle, R. "Spacetime Theory as Physical Geometry." *Erkenntnis* 42, no. 3 (1995): 317-337.

Dongen, Jeroen van. "Einstein and the Kaluza-Klein Particle." *Studies in History and Philosophy of Modern Physics* 33B, no. 2 (2002): 185-210.

Ducheyne, Steffen. "The Argument(s) for Universal Gravitation." *Foundations of Science* 11, no. 4 (2006): 419-447.

Duff, M. J. "Kaluza-Klein Theory in Perspective." In *Oskar Klein Centenary: Symposium* , 22-35. Singapore; River Edge, NJ: World Scientific, 1995.

Dupré, John. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, Mass.: Harvard University Press, 1993.

Earman, John. "Laws, Symmetry, and Symmetry Breaking: Invariance, Conservation Principles, and Objectivity." *Philosophy of Science* 71, no. 5 (2004): 1227.

———. "Tracking Down Gauge: An Ode to the Constrained Hamiltonian Formalism." In *Symmetries in Physics: Philosophical Reflections*, edited by Katherine Brading and Elena Castellani. Cambridge, U.K.; New York: Cambridge University Press, 2003.

Earman, John. "Curie's Principle and Spontaneous Symmetry Breaking." *International Studies in the Philosophy of Science* 18, no. 2-3 (July, 2004): 173-198.

———. *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*. New York: Oxford University Press, 1995.

———. "Some Aspects of General Relativity and Geometrodynamics." *Journal of Philosophy* 69, (October, 1972): 634-647.

Earman, John and John Norton. "What Price Spacetime Substantivalism: The Hole Story." *British Journal for the Philosophy of Science* 38, (December, 1987): 515-525.

Earman, John, Christopher Smeenk, and Christian Wüthrich. "Do the Laws of Physics Forbid the Operation of Time Machines?" *Synthese* 169, no. 1 (07/01, 2009): 91-124.

Eddington, Arthur Stanley. *Report on the Relativity Theory of Gravitation (to the Physical Society of London)*. 2d ed. London: Fleetway Press, Ltd., 1920.

Einstein, A. "Über den Gegenwärtigen Stand der Feld-Theorie." In *Festschrift zum 70. Geburtstag von Prof. Dr. A. Stodola*. Zurich: Orell Füssli, 1929.

———. "An Address Delivered in 1920 at the University of Leiden." .

———. "Hamiltonsches Prinzip und Allgemeine Relativitätstheorie." *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin)* (1916): 1111-1116.

Einstein, A. and M. Grossmann. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Vol. 621914.

Einstein, A. and P. G. Bergmann. "On a Generalization of Kaluza's Theory of Electricity." *Annals of Mathematics* 39, no. 3 (1938): 683.

Einstein, A. and J. Grommer. *Beweis der Nichtexistenz eines überall regulären zentrisch symmetrischen Feldes nach der Feldtheorie von Kaluza*. Mathematica Et Physica; Scripta Hierosolymitana, Edited by A. Einstein. Vol. 1. Jerusalem, Israel: Hebrew University, 1923.

Einstein, Albert. *Relativity: The Special and the General Theory; a Popular Exposition*. Bonanza Paperback. [Über die spezielle und die allgemeine Relativitätstheorie.]. 17th ed. New York: Crown Publishers, 1961.

———. *The Meaning of Relativity*. The Stafford Little Lectures. [Relativitätstheorie] . Translated by Edwin Plimpton Adams, Ernst G. Straus, Sonja Bargmann. 5th ed. Vol. 1921. Princeton: Princeton University Press, 1955.

———. *Out of My Later Years*. New York: Philosophical Library, 1949.

———. "Physik und Realität." *Franklin Institute Journal* 221, (1936): 313-337.

———. *The Principle of Relativity*. New York: Dover, 1923.

———. *Die Grundlage der Allgemeinen Relativitätstheorie*. Leipzig: Verlag von Johann Ambrosious Barth, 1916.

———. "Zum Gegenwärtigen Stande des Gravitationsproblems." *Physikalische Zeitschrift* 14, (December, 1913): 487-500.

Ekspong, Gösta. *The Oskar Klein Memorial Lectures*. Vol. 2. Singapore: World Scientific, 1994.

———. *The Oskar Klein Memorial Lectures*. Vol. 1. Singapore: World Scientific, 1991.

Endicott, Ronald P. "Post-Structuralist Angst. Critical Notice: John Bickle, Psychoneural Reduction: The New Wave." *Philosophy of Science* 68, no. 3 (September, 2001): 377-393.

———. "Collapse of the New Wave." *Journal of Philosophy* 95, no. 2 (February, 1998): 53-72.

Fabbri, Luca. "Taking Kaluza Seriously Leads to a Non-Gauge-Invariant Electromagnetic Theory in a Curved Space-Time." *Annales De La Fondation Louis De Broglie* 29, no. 4 (2004).

Felsager, Björn. *Geometry, Particles, and Fields*. New York: Springer, 1998.

Feynman, Richard Phillips, Robert B. Leighton, and Matthew L. Sands. *The Feynman Lectures on Physics*. Redwood City, Calif.: Addison-Wesley, 1989.

Field, Hartry H. *Science without Numbers: A Defence of Nominalism*. Princeton, N.J.: Princeton University Press, 1980.

Fock, V. E. R. "Über die Invariante Form der Wellen- und der Bewegungsgleichungen Für Einen Geladenen Massenpunkt." *Zeitschrift Für Physik A Hadrons and Nuclei* V39, no. 2 (02/01/, 1926): 226-232.

Forster, Malcolm R. "Unification, Explanation, and the Composition of Causes in Newtonian Mechanics." *Studies in History and Philosophy of Science* 19, (March, 1988): 55-101.

French, Steven. "A Model-Theoretic Account of Representation (Or, I Don't Know Much about Art...but I Know it Involves Isomorphism)." *Philosophy of Science* 70, no. 5 (December, 2003): 1472-1483.

Friedman, Michael. *Foundations of Space-Time Theories: Relativistic Physics and Philosophy of Science*. Princeton: Princeton University Press, 1983.

———. "Theoretical Explanation." In *Reduction, Time and Reality: Studies in the Philosophy of the Natural Sciences*, edited by Richard Healey. Cambridge, U.K.; New York: Cambridge University Press, 1981.

———. "Explanation and Scientific Understanding." *Journal of Philosophy* 71, no. 17 (January, 1974): 5-19.

Frigg, R. and N. Cartwright. "String Theory Under Scrutiny." *Physics World* 20, no. September (2006): 14-15.

Frisch, Mathias. *Inconsistency, Asymmetry, and Non-Locality: A Philosophical Investigation of Classical Electrodynamics*. Oxford Studies in the Philosophy of Science. Oxford ; New York: Oxford University Press, 2005.

Georgi, Howard. *Lie Algebras in Particle Physics*. Frontiers in Physics. 2nd ed. Vol. 54. Reading, Mass.: Perseus Books, 1999.

Geroch, R. "Electromagnetism as an Aspect of Geometry? Already Unified Field Theory—The Null Field Case." *Annals of Physics* 36, (1966): 147-187.

Giere, Ronald N. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press, 1988.

Gijsberg, Victor. "Why Unification is neither Necessary nor Sufficient for Explanation." *Philosophy of Science* 74, no. 4 (2007): 481.

Glashow, Sheldon L. "Towards a Unified Theory: Threads in a Tapestry." *Reviews of Modern Physics* 52, no. 3 (1980): 539.

Glennan, Stuart. "Rethinking Mechanistic Explanation." *Philosophy of Science* 69, (2002): S342-S353.

Glymour, Clark. *Theory and Evidence*. Princeton: Princeton University Press, 1980.

Godfrey-Smith, Peter. *Theory and Reality: An Introduction to the Philosophy of Science*. Science and its Conceptual Foundations. Chicago: University of Chicago Press, 2003.

Goenner, Hubert F. M. "On the History of Unified Field Theories." *Living Reviews in Relativity* 7, no. 2 (2004).

Goodwin, William. "Scientific Understanding After the Ingold Revolution in Organic Chemistry." *Philosophy of Science* 74, no. 3 (July, 2007): 386-408.

Gordon Belot. "Understanding Electromagnetism." *The British Journal for the Philosophy of Science* 49, no. 4 (Dec., 1998): 531-555.

Green, Michael. "Superstrings." *Scientific American* 255, no. 3 (1986).

Greene, B. *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*. 1st ed. New York: W. W. Norton, 1999.

Grünbaum, Adolf. "Geometrodynamics and Ontology." *Journal of Philosophy* 70, no. 6 (December, 1973): 775-800.

Hacking, Ian. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge, Cambridgeshire ; New York: Cambridge University Press, 1983.

Halonen, Ilpo and Jaakko Hintikka. "Unification—it's Magnificent but is it Explanation?" *Synthese* 120, no. 1 (1999): 27-47.

Hartle, J. B. *Gravity : An Introduction to Einstein's General Relativity*. San Francisco: Addison-Wesley, 2003.

Hausman, Daniel and James Woodward. "Manipulation and the Causal Markov Condition." *Philosophy of Science* 71, no. 5 (December, 2004): 846-856.

Healey, Richard. *Gauging what's Real: The Conceptual Foundations of Contemporary Gauge Theories*. Oxford ; New York: Oxford University Press, 2007.

———. "Gauge Theories and Holisms." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 35, no. 4 (12, 2004): 619-642.

———. "On the Reality of Gauge Potentials." *Philosophy of Science* 68, no. 4 (December, 2001): 432-455.

Heath, Thomas Little. *Aristarchus of Samos, the Ancient Copernicus*. New York: Dover Publications, 1913.

Hempel, C. G. and P. Oppenheim. "Studies in the Logic of Explanation." *Philosophy of Science* 15, (April, 1948): 135-175.

Hempel, Carl Gustav. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press, 1965.

Hendry, Robin Findlay and Paul Needham. "Le Poidevin on the Reduction of Chemistry." *British Journal for the Philosophy of Science* 58, no. 2 (June 1, 2007): 339-353.

Henry, John. "A Cambridge Platonist's Materialism: Henry More and the Concept of Soul." *Journal of the Warburg and Courtauld Institutes* 49, (1986): 172-195.

Hermann, Robert. *Yang-Mills, Kaluza-Klein, and the Einstein Program*. Vol. XIX. Brookline, Mass.: Math Sci Press, 1978.

Hilpinen, Risto (eds). *Rationality in Science, Philosophy of Science, Proceedings of the PSA meeting, V21,* 1980.

Hinton, Charles Howard. "The Fourth Dimension." (1904 (1912)).

Hon, Giora and Bernard R. Goldstein. "Unpacking 'for Reasons of Symmetry': Two Categories of Symmetry Arguments." *Philosophy of Science* 73, no. 4 (2006): 419-439.

Hooker, C. A. "Towards a General Theory of Reduction, Part I: Historical and Scientific Setting." *Dialogue: Canadian Philosophical Review* 20, (March, 1981): 38-59.

Horgan, J. *The End of Science. Facing the Limits of Knowledge in the Twilight*. London: Abacus, 1996.

Howard, Don, ed. “Einstein’s Philosophy of Science”, in *Stanford Encyclopedia of Philosophy*, accessed 2009, June.

Hudson, Hud. *The Metaphysics of Hyperspace*. Oxford; New York: Oxford University Press, 2005.

Huggett, Nick and Robert Weingard. "Exposing the Machinery of Infinite Renormalization." *Philosophy of Science* 63, no. 3 (09/02, 1996): S159.

Janssen, M. "Of Pots and Holes: Einstein's Bumpy Road to General Relativity." *Annalen Der Physik.* 14, (2005): 58.

Johnson, Monte Ransome. *Aristotle on Teleology*. Oxford Aristotle Studies. Oxford; New York: Clarendon Press; Oxford University Press, 2005.

Jones, Todd. "How the Unification Theory of Explanation Escapes Asymmetry Problems." *Erkenntnis: An International Journal of Analytic Philosophy* 43, no. 2 (September, 1995a): 229-240.

———. "Reductionism and the Unification Theory of Explanation." *Philosophy of Science* 62, no. 1 (March, 1995b): 21-30.

Kafatos, F. C. and T. Eisner. "Unification in the Century of Biology." *Science (New York, N.Y.)* 303, no. 5662 (Feb 27, 2004): 1257.

Kaku, Michio. *Hyperspace: A Scientific Odyssey through Parallel Universes, Time Warps, and the 10th Dimension*. Oxford; New York: Oxford University Press, 1994.

Kalinowski, M. W. "The Program of Geometrization of Physics: Some Philosophical Remarks." *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science* 77, (November, 1988): 129-138.

Kaluza, Theodor. "On the Unity Problem of Physics." In *Modern Kaluza-Klein Theories*, edited by Thomas Appelquist, Alan Chodos and Peter G. O. Freund, 61-691986 [1921].

———. "Zum Unitätproblem Der Physik." *Sitzungsberichte Der Königlich Preußischen Akademie Der Wissenschaften Zu Berlin* (1921): 966-972.

Kellert, Stephen H., Helen E. Longino, and C. Kenneth Waters. *Scientific Pluralism*. Minneapolis, MN: University of Minnesota Press, 2006.

Kerner, R. "Generalization of the Kaluza-Klein Theory for an Arbitrary Non-Abelian Gauge Theory." *Annales De La Institut Henri Poincare* 9, (1968): 143.

Kilmister, C. W. *Schrödinger, Centenary Celebration of a Polymath*. Cambridge Cambridgeshire ; London: Cambridge University Press, 1987.

Kiritsis, Elias. *String Theory in a Nutshell*. Princeton: Princeton University Press, 2007.

Kitcher, P. "Unification as a Regulative Ideal." *Perspectives on Science* 7, (1999): 337-348.

Kitcher, Philip. "Explanatory Unification and the Causal Structure of the World." Chap. 11, In *Scientific Explanation*, edited by Wesley Salmon. Vol. XIII. Minneapolis: University of Minnesota Press, 1989.

———. "Explanatory Unification." *Philosophy of Science* 48, (December, 1981): 507-531.

———. "Explanation, Conjunction, and Unification." *Journal of Philosophy* 73, no. 22 (April, 1976): 207-212.

Klein, Oskar. "Zur Fünfdimensionale Darstellung der Relativitätstheorie." *Zeitschrift für Physik* 46, no. 3 (1928): 188-208.

———. "Quantentheorie und Fünfdimensionale Relativitätstheorie." *Zeitschrift für Physik A Hadrons and Nuclei* V37, no. 12 (12/01/, 1926): 895-906.

Klein, Oskar. "Quantum Theory and Five Dimensional Theory of Relativity." In *Modern Kaluza-Klein Theories*, edited by Thomas Appelquist, Alan Chodos and Peter G. O. Freund. Menlo Park: Addison Wesley, 1981 [1926].

———. "The Atomicity of Electricity as a Quantum Theory Law." *Nature* 118, no. 2971 (1926): 516.

Kneale, W. C. *Probability and Induction*. 2nd ed. Oxford: Clarendon Press, 1952.

Kosso, P. "Symmetry, Objectivity and Design." In *Symmetries in Physics: Philosophical Reflections*, edited by K. Brading and E. Castellani, 413-465. Cambridge: Cambridge University Press, 2003.

Kosso, P. "The Empirical Status of Symmetries in Physics." *British Journal for the Philosophy of Science* 51, (2000): 81-98.

Kosso, Peter. "Science and Objectivity." *The Journal of Philosophy* 86, no. 5 (May, 1989): 245-257.

Kox, A. J. *The Collected Papers of Albert Einstein. the Berlin Years: Writings, 1914-1917*. Vol. 6. Princeton: Princeton University Press, 1987.

Kragh, Helge. "[the] Equation with the Many Fathers. the Klein–Gordon Equation in 1926 " *American Journal of Physics* 52, no. 11 (1984): 1024.

Kramers, H. A. "On the Application of Einstein's Theory of Gravitation to a Stationary Field of Gravitation." *Proceedings of the Royal Netherlands Academy of Arts and Sciences* 23, no. 7 (1922): 1052-1073.

Kukla, An. "Unification as a Goal for Psychology." *American Psychologist* 47, no. 8 (1992): 1054.

Kukla, André. "Scientific Realism and Theoretical Unification." *Analysis* 55, no. 4 (October, 1995): 230-238.

Lange, Marc. "Bayesianism and Unification: A Reply to Wayne Myrvold." *Philosophy of Science* 71, no. 2 (April, 2004): 205-215.

———. *Natural Laws in Scientific Practice*. Oxford ; New York: Oxford University Press, 2000.

Leeds, Stephen. "Gauges: Aharonov, Bohm, Yang, Healey." *Philosophy of Science* 66, no. 4 (Dec., 1999): 606-627.

Lehmkuhl, Dennis. "Is Spacetime a Gravitational Field?" In *Ontology of Spacetime*, edited by Dennis Dieks. Vol. 2, 83-110. Amsterdam: Elsevier, 2008.

Leibniz, Gottfried Wilhelm. *Selections*, edited by Philip P. Wiener. New York: Scribner, 1951.

Lewis, D. "Causal Explanation." *Philosophical Papers* 2, (1986): 214–240.

Lichnerowicz, André. *Théories Relativistes de la Gravitation et de l'Électromagnétisme; Relativité Générale et Théories Unitaires*. Collection d'Ouvrages de Mathématiques à l'Usage des Physiciens. Paris: Masson, 1955.

Lipton, Peter. *Inference to the Best Explanation*. Second Ed. ed. London; New York: Routledge, 2004.

London, F. "Quantenmechanische Deutung der Theorie von Weyl (Quantum-Mechanical Interpretation of Weyl’s Theory)." *Zeitschrift für Physik* 42, (1927): 375-389.

Lorentz, H. A., A. Einstein, H. Minkowski, and H. Weyl. *The Principle of Relativity; a Collection of Original Memoirs on the Special and General Theory of Relativity* .

Translated by W. l. Perrett and G. B. Jeffery, edited by A. Sommerfeld. New York: Dover, 1952.

Lyre, Holger. "Relativity and Fields--the Principles of Gauging." *Philosophy of Science.* 68, no. 3 (2001): S371.

Lyre, Holger and Tim Oliver Eynck. "Curve it, Gauge it, Or Leave it? Practical Underdetermination in Gravitational Theories." *Journal for General Philosophy of Science* 34, no. 2 (2003): 277-303.

Mach, E. *The Analysis of Sensations, and the Relation of the Physical to the Psychical* . Translated by C. M. Williams. New York: Dover Publications Inc., 1959.

Mach, Ernst. *The Science of Mechanics; a Critical and Historical Exposition of its Principles*. Translated by Thomas McCormack and J. Ambrose Flemming. Chicago: Open Court, 1893.

Magnani, Lorenzo. *Philosophy and Geometry: Theoretical and Historical Issues*. Western Ontario Series in Philosophy of Science. Vol. 66. Dordrecht; Boston: Kluwer Academic Publishers, 2001.

Majer, U. and Tilman Sauer. "Hilbert's 'World Equations' and His Vision of a Unified Science." In *The Universe of General Relativity*, edited by A. J. Kox and J. Eisenstaedt. Boston; Basel; Berlin: Birkhauser, 2005.

Malament, David. "Classical Relativity Theory." In *Philosophy of Physics*, edited by Jeremy Butterfield and John Earman. 1st ed., 1434-1453. Amsterdam; Boston: Elsevier/North-Holland, 2007.

Maudlin, Tim. "On the Unification of Physics." *Journal of Philosophy* 93, no. 3 (March, 1996): 129-144.

Maxwell (Clerk), J. "On Physical Lines of Force." *Philosophical Magazine* 21, (1861).

Maxwell, James Clerk. *A Dynamical Theory of the Electromagnetic Field.* (London): 1865.

Maxwell, Nicholas. *Non-Empirical Requirements Scientific Theories Must Satisfy: Simplicity, Unification, Explanation, Beauty*2004, http://philsci-archive.pitt.edu/archive/00001759/ .

———. *The Comprehensibility of the Universe: A New Conception of Science*. Second ed.1998.

McCormmach, Russell. "H. A. Lorentz and the Electromagnetic View of Nature." *Isis* 61, no. 4 (Winter, 1970): 459-497.

Melia, J. "Weaseling Away the Indispensability Argument." *Mind* 109, no. 435 (2000): 455.

Meyerson, Emile. *Explanation in the Sciences*. Boston Studies in the Philosophy of Science. [De l'explication dans les sciences.]. Vol. 128. Dordrecht ; Boston: Kluwer Academic Publishers, 1991.

———. *The Relativistic Deduction: Epistemological Implications of the Theory of Relativity*. Boston Studies in the Philosophy of Science [Déduction relativiste.]. Translated by A. David and Mary-Alice Sipfle. Vol. 83. Dordrecht; Boston; Hingham, MA: D. Reidel; Kluwer Academic (USA), 1985 (1925).

Meyerson, Emile,. *Identity & Reality*. New York: Dover Publications, 1962.

Minkowski, H. "Die Grundgleichungen Für Die Elektromagnetischen Vorgänge in Bewegten Körpern." *Mathematische Annalen. Akademie Der Wissenschaften in Göttingen, Klasse M.P.* 68, no. 4 (1910): 472-525.

———. "Raum Und Zeit." *Physikalische Zeitschrift* 20, (1909): 104-111.

Misner, C. W. and J. A. Wheeler. "Gravitation, Electromagnetism, Unquantized Charge, and Mass as Properties of Curved Empty Space " *Annals of Physics* 2, (1957): 525-603.

Misner, Charles W., Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. San Francisco: W. H. Freeman, 1973.

Moriyasu, K. *An Elementary Primer for Gauge Theory*. Singapore: World Scientific, 1983.

Morrison, Margaret. "Unification, Explanation and Explaining Unity: The Fisher-Wright Controversy." *The British Journal for the Philosophy of Science* 57, no. 1 (March 1, 2006): 233-245.

———. *Unifying Scientific Theories: Physical Concepts and Mathematical Structures*-Cambridge University Press, 2000.

———. "Unified Theories and Disparate Things." *Proceedings of the Biennial Meetings of the Philosophy of Science Association* 2, (1995): 365-373.

———. "A Study in Theory Unification: The Case of Maxwell's Electromagnetic Theory." *Studies in History and Philosophy of Science* 23, no. 1 (March, 1992): 103-145.

———. "Unification, Realism and Inference." *British Journal for the Philosophy of Science* (September, 1990): 305-332.

Muller, F. A. "Margaret Morrison, Critical Discussion of Unifying Scientific Theories. Physical Concepts and Mathematical Structures." *Erkenntnis* 55, (2001): 132-143.

Muntean, Ioan. "Mechanisms of Unification in Kaluza-Klein Theory." In *Ontology of Spacetime*, edited by Dennis Dieks. Vol. 2, 279-299. Amsterdam: Elsevier, 2008.

———. "The “Field” Interpretation of Newton’s Philosophy of Nature." 2005 (not published).

Myrvold, Wayne C. "A Bayesian Account of the Virtue of Unification." *Philosophy of Science* 70, no. 2 (April, 2003): 399-423.

Nahin, Paul J. *Time Machines: Time Travel in Physics, Metaphysics, and Science Fiction*. New York, N.Y.: American Institute of Physics, 1993.

Nerlich, Graham. *What Spacetime Explains: Metaphysical Essays on Space and Time*. Cambridge ; New York: Cambridge University Press, 1994.

Nicolai, K., K. Peeters, and M. Zamaklar. "Topical Review: Loop Quantum Gravity: An Outside View." *Classical and Quantum Gravity* 22, (oct, 2005): 193.

Nolan, Daniel. "Quantitative Parsimony." *The British Journal for the Philosophy of Science* 48, no. 3 (September 1, 1997): 329-343.

Norton, J. D. "A Material Theory of Induction." *Philosophy of Science* 70, (2003): 647-670.

———. "Einstein, Nordström and the Early Demise of Scalar, Lorentz-Covariant Theories of Gravitation." *Archive for History of Exact Sciences* 45, no. 1 (1992): 17-94.

Norton, John. "What was Einstein's Principle of Equivalence?" *Studies in History and Philosophy of Science Part A*, 16, no. 3 (9, 1985): 203-246.

Norton, John. "How Einstein found His Field Equations: 1912-1915." *Historical Studies in the Physical Sciences* 14, (1984).

Oppenheim, P. and H. Putnam. "Unity of Science as a Working Hypothesis." In *Minnesota Studies in the Philosophy of Science*, edited by H. Feigl, G. Maxwell and M. Scriven. Vol. 1. Minnesota: Minnesota University Press, 1958.

O'Raifeartaigh, L. *The Dawning of Gauge Theory*. Princeton: Princeton University Press, 1997.

O'Raifeartaigh, L. and N. Straumann. "Gauge Theory: Historical Origins and some Modern Sevelopments." *Reviews of Modern Physics* 72, no. 1 (January, 2000): 1-23.

Ortìn, Tomás. *Gravity and Strings*. Cambridge: Cambridge University Press, 2004.

Overduin, J. M. and P. S. Wesson. "Kaluza-Klein Gravity." *Physics Reports* 283, no. 5-6 (1997): 303.

Pais, Abraham. *The Genius of Science: A Portrait Gallery*. Oxford; New York: Oxford University Press, 2000.

———. *"Subtle is the Lord...": The Science and the Life of Albert Einstein*Oxford University Press, 1982.

Pasini, Antonello. "A Conceptual Introduction to the Kaluza-Klein Theory." *European Journal of Physics* 9, (1988): 289-296.

Pauli, Wolfgang. *Theory of Relativity* [Relativitätstheorie.]. New York: Pergamon Press, 1958.

Pearson, K. "On Lines and Planes of Closest Fit to Systems of Points in Space." *Philosophical Magazine* 2, no. 6 (1901): 559-572.

Penrose, Roger. *The Road to Reality: A Complete Guide to the Laws of the Universe*. 1st American ed. New York: A.A. Knopf, 2005.

Petkov, Vesselin. *Relativity and the Dimensionality of the World*. Dordrecht, The Netherlands: Springer, 2007.

Plutynski, Anya. "Explanatory Unification and the Early Synthesis." *The British Journal for the Philosophy of Science* 56, no. 3 (September 1, 2005): 595-609.

Poincaré, Henri. *La Science et l'Hypothèse*. Bibliothèque de Philosophie Scientifique. Paris: Ernest Flammarion, 1920.

———. "Sur La Dynamique De l'Électron." *Rendiconti Del Circolo Matematico Di Palermo* 21, (1906): 129-175.

Polchinski, Joseph Gerard. *String Theory: An Introduction to the Bosonic String*. Cambridge Monographs on Mathematical Physics. Pbk ed. Cambridge ; New York: Cambridge University Press, 2005.

Popper, Karl R. *Conjectures and Refutations; the Growth of Scientific Knowledge*. New York: Basic Books, 1962.

Psillos, Stathis. *Causation and Explanation*. Central Problems of Philosophy. Montréal: McGill-Queen's University Press, 2002.

Putnam, Hilary. *Mind, Language, and Reality*. Philosophical Papers. 1st paperback ed. Vol. 2. Cambridge ; New York: Cambridge University Press, 1979.

Quine, Willard van Orman. "On Simple Theories of a Complex World." *Synthese* 15, no. 1 (1963): 103.

Quine, WV. "On Empirically Equivalent Systems of the World." *Erkenntnis* 9, no. 3 (1975): 313-328.

Railton, Peter. "Probability, Explanation, and Information." *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science* 48, (August, 1981): 233-256.

Randall, Lisa. *Warped Passages : Unraveling the Mysteries of the Universe's Hidden Dimensions*. 1st ed. New York: Ecco, 2005.

———. "Large Mass Hierarchy from a Small Extra Dimension " *Physical Review Letters* 83, no. 17 (1999): 3370.

Redhead, M. L. G. "Symmetry in Intertheory Relations." *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science* 32, (November-December, 1975): 77-112.

Redhead, Michael. "The Interpretation of Gauge Symmetry." In *Symmetries in Physics: Philosophical Reflections,* edited by Elena Castellani and Katherine Brading. Cambridge U.K.; New York: Cambridge University Press, 2003.

Redhead, Michael, Jeremy Butterfield, and Constantine Pagonis. "From Physics to Philosophy."Cambridge University Press, 1999.

Redhead, Michael and Paul Teller. "Particle Labels and the Theory of Indistinguishable Particles in Quantum Mechanics." *The British Journal for the Philosophy of Science* 43, no. 2 (1992): 201-218.

Reichenbach, Hans. *The Philosophy of Space and Time* [Philosophie der Raum-Zeit-Lehre.]. New York: Dover Publications, 1958 (1928).

Reissner, H. "Über Die Eigengravitation Des Elektrischen Feldes Nach Der Einsteinschen Theorie." *Annalen Der Physik* 355, no. 9 (1916): 106.

Renn, Jürgen, ed. *Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy*. The Genesis of General Relativity. Vol. 3. Dordrecht: Springer, 2007.

Renn, Jürgen and Matthias Schemmel, eds. *The Promise of Mathematics. Gravitation in the Twilight of Classical Physics*. The Genesis of General Relativity. Vol. 4. Dordrecht: Springer, 2007.

Riemann, Georg Friedrich Bernhard. *Gesammelte Mathematische Werke Und Wissenschaftlicher Nachlass*. 2. Aufl. bearb. von Heinrich Weber ed. New York: Dover Publications, 1953.

Rindler, Wolfgang. *Relativity: Special, General, and Cosmological*. 2nd ed. Oxford ; New York: Oxford University Press, 2006.

Robbin, Tony. *Shadows of Reality: The Fourth Dimension in Relativity, Cubism, and Modern Thought*. New Haven: Yale University Press, 2006.

Rothman, Tony and Stephen Boughn. "Can Gravitons be Detected?" *Foundations of Physics* 36, no. 12 (12/01, 2006): 1801-1825.

Roush, Sherrilyn. "Testability and the Unity of Science." *Journal of Philosophy* 101, no. 11 (November, 2004): 555-573.

Rovelli, Carlo. "Halfway through the Woods: Contemporary Research on Space and Time." In *The Cosmos of Science: Essays of Exploration,* edited by John Earman and John D. Norton. Pittsburgh, Pa.; Konstanz: University of Pittsburgh Press ; Universitatsverlag Konstanz, 1997.

Rovelli, Carlo. *Quantum Gravity*. Cambridge, UK; New York: Cambridge University Press, 2004.

Rucker, Rudy. *Geometry, Relativity, and the Fourth Dimension*. New York: Dover Publications, 1977.

Ryckman, Thomas. *The Reign of Relativity: Philosophy in Physics, 1915-1925*Oxford University Press, 2005.

Sabbata, Venzo De and Ernst Schmutzer. "Unified Field Theories of More than 4 Dimensions Including Exact Solutions." Erice, Trapani, Sicily, World Scientific, 1982.

Salam, A. and J. Strathdee. "On Kaluza-Klein Theory." *Annals of Physics* 141, no. 2 (1982): 316-352.

Salmon, Wesley. *Scientific Explanation*. Minnesota Studies in the Philosophy of Science. Vol. XI. Minneapolis: University of Minnesota Press, 1989.

Salmon, Wesley C. *Four Decades of Scientific Explanation*. 1st University of Pittsburgh Press pbk. ed. Pittsburgh: University of Pittsburgh Press, 2006.

———. *Causality and Explanation*. New York: Oxford University Press, 1998.

———. "Scientific Explanation: Causation and Unification." *Critica: Revista Hispanoamericana De Filosofia* 22, no. 66 (December, 1990): 3-24.

———. *Scientific Explanation and the Causal Structure of the World*. Princeton, N.J.: Princeton University Press, 1984.

Schilpp, Paul Arthur and Albert Einstein. *Albert Einstein, Philosopher-Scientist*. The Library of Living Philosophers. 1st ed. Vol. 7. Evanston, Ill.: Library of Living Philosophers, 1949.

Schlick, M. *Allgemeine Erkenntnisslehre*. Berlin: Springer, 1918.

Scholz, Erhard, ed. *Hermann Weyl's Raum-Zeit-Materie and a General Introduction to His Scientific Work*. DMV Seminar. Vol. 30. Basel ; Boston: Birkhauser Verlag, 2001.

Schrödinger, Erwin. "Über Eine Bemerkenswerte Eigenschaft Der Quantenbahnen Eines Einzelnen Elektrons." *Zeitschrift Für Physik A Hadrons and Nuclei* V12, no. 1 (1923): 13-23.

Schupbach, Jonah N. "On a Bayesian Analysis of the Virtue of Unification." *Philosophy of Science* 72, no. 4 (October, 2005): 594-607.

Schurz, Gerhard. "Explanation as Unification." *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science* 120, no. 1 (1999): 95-114.

Schurz, Gerhard and Karel Lambert. "Outline of a Theory of Scientific Understanding." *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science* 101, no. 1 (October, 1994): 65-120.

Seahra, S. S. and P. S. Wesson. "Application of the Campbell-Magaard Theorem to Higher-Dimensional Physics." *Classical and Quantum Gravity* 20, no. 7 (2003): 1321-1339.

Shapiro, Stewart. *Thinking about Mathematics: The Philosophy of Mathematics*. New York: Oxford University Press, 2000.

Sider, Theodore. *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford; New York: Clarendon Press; Oxford University Press, 2001.

Sklar, Lawrence. "Dappled Theories in a Uniform World." *Philosophy of Science* 70, no. 2 (April, 2003): 424-441.

———. *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge England ; New York: Cambridge University Press, 1993.

———. *Space, Time and Spacetime*. Berkeley: University of California Press, 1974.

Smeenk, Chris. "The Elusive Higgs Mechanism." *Philosophy of Science* 73, no. 5 (December, 2006): 487-499.

Smith, Peter. "Modest Reductions and the Unity of Science." In *Reduction, Explanation and Realism*, edited by D. Charles and K. Lennon. Oxford: Clarendon Press; Oxford University Press, 1992.

Smolin, Lee. *The Trouble with Physics: The Rise of String Theory, the Fall of a Science, and what Comes Next*. Boston, New York: Houghton Mifflin, 2006.

———. "The Case for Background Independence." (2005), hep-th/0507235.

Speiser, Andreas and Hermann Weyl. *Gesammelte Abhandlungen Von Hermann Minkowski*. Wiesbaden: B. G. Teubner, 1911.

Spirtes, Peter. "From Probability to Causality." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* (October, 1991): 1-36.

Stein, Howard. "On the Notion of Field in Newton, Maxwell and Beyond." In *Historical and Philosophical Perspectives of Science*, edited by Roger H. Stuewer. Vol. 1, 384. New York: Gordon and Breach, 1989; 1970.

Stein, Howard. "Graves on the Philosophy of Physics." *Journal of Philosophy* 69, (October, 1972): 621-634.

Steinhardt, Paul J. and Neil Turok. *Endless Universe: Beyond the Big Bang*. New York: Doubleday, 2007.

Stöltzner, Michael. "The Principle of Least Action as the Logical Empiricist's Shibboleth." *Studies in History and Philosophy of Modern Physics* 34, pt. B, no. 2 (June, 2003): 285-318.

Strevens, Michael. "The Causal and Unification Approaches to Explanation Unified-Causally." *Nous* 38, no. 1 (2004): 154-176.

Strevens, Michael. *Depth: An Account of Scientific Explanation*. Cambridge, Mass.: Harvard University Press, 2008.

Suppe, F. *The Structure of Scientific Theories*. Champaign: University of Illinois Press, 1977.

Susskind, Leonard. *Cosmic Landscape: String Theory and the Illusion of Intelligent Design*. 1st ed. New York: Little, Brown and Co., 2005.

Teller, Paul. "How we Dapple the World." *Philosophy of Science* 71, no. 4 (October, 2004): 425-447.

Thorne, Kip S. *Black Holes and Time Warps: Einstein's Outrageous Legacy*. New York: W.W. Norton, 1994.

Tonnelat, M. A. *Les Théories Unitaires De l'Électromagnétisme Et De La Gravitation*. Paris: Gauthier-Villars, 1965.

Torretti, Roberto. *Philosophy of Geometry from Riemann to Poincaré*. Pallas Paperback. Dordrecht, Holland; Hingham, MA: D. Reidel ; Kluwer Academic, 1984.

Trautman, A. "Fibre Bundles Associated with Space-Time." *Rep.Math.Phys* 1, (1970): 29.

———. "Foundations and Current Problems of General Relativity." In *Lectures on General Relativity*, edited by S. Deser and K. W. Ford. Upper Saddle Rive, New Jersey: Prentice Hall, 1965.

van Dongen, J. "Einstein's Unification: General Relativity and the Quest for Mathematical Naturalness."University of Amsterdam, 2002.

Van Fraassen, Bas C. *Laws and Symmetry*. Oxford ; New York: Oxford University Press, 1989.

———. *The Scientific Image*. Clarendon Library of Logic and Philosophy. Oxford; New York: Clarendon Press; Oxford University Press, 1980.

———. "The Pragmatics of Explanation." *American Philosophical Quarterly* 14, (April, 1977): 143-150.

Veblen, Oswald and Banesh Hoffmann. "Projective Relativity." *Physical Review* 36, no. 5 (Sep, 1930): 810-822.

Vizgin, Vladimir. *Unified Field Theories in the First Third of the 20th Century* . Translated by Julian B. Barbour. Basel ed. Birkhauser Verlag, 1994.

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

Watkins, John W. N. *Science and Scepticism*. Princeton, N.J.: Princeton University Press, 1984.

Wayne, Andrew. "Theoretical Unity: The Case of the Standard Model." *Perspectives on Science* 4, no. 4 (Winter96, 1996): 391.

Weber, Erik. "Unification and Explanation: A Comment on Halonen and Hintikka, and Schurz." *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science* 131, no. 1 (April, 2002): 145-154.

Weinberg, S. *Dreams of a Final Theory: The Scientist's Search for the Ultimate Laws of Nature*. New York: Vintage, 1992.

———. "Lecture on the Applicability of Mathematics." *Notices of the American Mathematical Society* 33, no. 5 (1986): 725-728.

Weinberg, Steven. "Viewpoints on String Theory: The Elegant Universe by B. Greene." NOVA & PBS. http://www.pbs.org/wgbh/nova/elegant/view-weinberg.html (accessed 7/2/2008, 2008).

———. "Conceptual Foundations of the Unified Theory of Weak and Electromagnetic Interactions." *Reviews of Modern Physics* 52, (1980): 515-523.

Weingard, R. "A Philosopher Looks at String Theory." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1988, (1988): 95-106.

Weingard, Robert. "On Two Goals of Unification in Physics.".

Weinstein, Steven. "Gravity and Gauge Theory." *Philosophy of Science* 66, no. 3 Supplement (1999): S146-S155.

Wesson, Paul S. *Space-Time-Matter: Modern Higher-Dimensional Cosmology*. 2nd ed. Singapore; Hackensack, N.J.: World Scientific, 2007.

———. "In Defense of Campbell's Theorem as a Frame for New Physics." (2005).

Weyl, H. "Gravitation Und Elektrizitat." *Sitzungsber.Preuss.Akad.Wiss* 26, (1918): 465–478.

Weyl, Hermann. *Raum, Zeit, Materie*. 6., unveränd. Aufl ed. Berlin, Heidelberg, New York: Springer, 1970.

———. *Symmetry*. Princeton: Princeton University Press, 1952.

Wheeler, John Archibald. *Geometrodynamics*. Società Italiana Di Fisica. Questioni Di Fisica Moderna. Vol. 1. New York: Academic Press, 1962.

Wigner, Eugene Paul. *Symmetries and Reflections; Scientific Essays of Eugene P. Wigner*. Bloomington: Indiana University Press, 1967.

Witten, E. "Mass Hierarchies in Supersymmetric Theories." *Physics Letters B* 105, no. 4 (1981a): 267-271.

———. "Search for a Realistic Kaluza-Klein Theory." *Nucl.Phys.B* 186, (1981b): 412.

Woit, Peter. *Not Even Wrong: The Failure of String Theory and the Search for Unity in Physical Law*. New York: Basic Books, 2006.

Wolfson, Paul and James Woodward. "Scientific Explanation and Sklar's Views of Space and Time." *Philosophy of Science* 46, no. 2 (Jun., 1979): 287-294.

Woodward, James. *Making Things Happen*. Oxford Studies in Philosophy of Science. New York: Oxford University Press, 2003.

———. "A Theory of Singular Causal Explanation." *Erkenntnis: An International Journal of Analytic Philosophy* 21, (November, 1984): 231-262.

Wünsch, Daniela. *Der Erfinder Der 5. Dimension : Theodor Kaluza : Leben Und Werk*. Göttingen: Termessos, 2007.

Wünsch, Daniela. "Einstein, Kaluza and the Fifth Dimension." In *The Universe of General Relativity*, edited by A. J. Kox and J. Eisenstaedt. Boston; Basel; Berlin: Birkhauser, 2005.

Wünsch, Daniela and Hubert Goenner. *Kaluza's and Klein's Contributions to the Kaluza-Klein Theory*Max Planck Institute for the History of Science, 2005.

Wüthrich, Christian. "Approaching the Planck Scale from a Generally Relativistic Point of View: A Philosophical Appraisal of Loop Quantum Gravity." Ph D, University of Pittsburgh, 2006.

Wüthrich, Christian. "To Quantize Or Not to Quantize: Fact and Folklore in Quantum Gravity." *Philosophy of Science* 72, (2005): 777-788.

Zahar, Elie. *Why Science Needs Metaphysics: A Plea for Structural Realism*. Chicago: Open Court, 2007.

———. "Einstein, Meyerson and the Role of Mathematics in Physical Discovery." *The British Journal for the Philosophy of Science* 31, no. 1 (Mar., 1980): 1-43.

Zwiebach, Barton. *A First Course in String Theory*. New York: Cambridge University Press, 2004.