UNIL | Université de Lausanne

Faculté des lettres

FACULTÉ DES LETTRES

SECTION DE PHILOSOPHIE

# On the Applicability of Mathematics – Philosophical and Historical Perspectives

THÈSE DE DOCTORAT

présentée à la

Faculté des lettres
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès lettres

par

**Tim Räz**

Directeur de thèse

Michael-Andreas Esfeld

LAUSANNE

2013

UNIL | Université de Lausanne
Faculté des lettres

# IMPRIMATUR

Le Décanat de la Faculté des lettres, sur le rapport d'une commission composée de :

**Directeur de thèse :**

| | |
|---|---|
| Monsieur Michael-Andreas Esfeld | Professeur, Faculté des lettres, Université de Lausanne |

**Membres du jury :**

| | |
|---|---|
| Monsieur Claus Beisbart | Professeur de philosophie, Université de Berne |
| Monsieur Hannes Leitgeb | Professeur de philosophie, Ludwig-Maximilians-Université München, Allemagne |

autorise l'impression de la thèse de doctorat de

## MONSIEUR TIM RÄZ

intitulée

## On the Applicability of Mathematics – Philosophical and Historical Perspectives.

sans se prononcer sur les opinions du candidat / de la candidate.

La Faculté des lettres, conformément à son règlement, ne décerne aucune mention.

Lausanne, le 2 décembre 2013

François Rosset
Doyen de la Faculté des lettres

# Abstract

The present thesis is a contribution to the debate on the applicability of mathematics; it examines the interplay between mathematics and the world, using historical case studies.

The first part of the thesis consists of four small case studies. In chapter 1, I criticize "ante rem structuralism", proposed by Stewart Shapiro, by showing that his so-called "finite cardinal structures" are in conflict with mathematical practice. In chapter 2, I discuss Leonhard Euler's solution to the Königsberg bridges problem. I propose interpreting Euler's solution both as an explanation within mathematics and as a scientific explanation. I put the insights from the historical case to work against recent philosophical accounts of the Königsberg case. In chapter 3, I analyze the predator-prey model, proposed by Lotka and Volterra. I extract some interesting philosophical lessons from Volterra's original account of the model, such as: Volterra's remarks on mathematical methodology; the relation between mathematics and idealization in the construction of the model; some relevant details in the derivation of the Third Law, and; notions of intervention that are motivated by one of Volterra's main mathematical tools, phase spaces. In chapter 4, I discuss scientific and mathematical attempts to explain the structure of the bee's honeycomb. In the first part, I discuss a candidate explanation, based on the mathematical Honeycomb Conjecture, presented in Lyon and Colyvan (2008). I argue that this explanation is not scientifically adequate. In the second part, I discuss other mathematical, physical and biological studies that could contribute to an explanation of the bee's honeycomb. The upshot is that most of the relevant mathematics is not yet sufficiently understood, and there is also an ongoing debate as to the biological details of the construction of the bee's honeycomb.

The second part of the thesis is a bigger case study from physics: the genesis of GR. Chapter 5 is a short introduction to the history, physics and mathematics that is relevant to the genesis of general relativity (GR). Chapter 6 discusses the historical question as to what Marcel Grossmann contributed to the genesis of GR. I will examine the so-called "Entwurf" paper, an important joint publication by Einstein and Grossmann, containing the first tensorial formulation of GR. By comparing Grossmann's part with the mathematical theories he used, we can gain a better understanding of

what is involved in the first steps of assimilating a mathematical theory to a physical question. In chapter 7, I introduce, and discuss, a recent account of the applicability of mathematics to the world, the Inferential Conception (IC), proposed by Bueno and Colyvan (2011). I give a short exposition of the IC, offer some critical remarks on the account, discuss potential philosophical objections, and I propose some extensions of the IC. In chapter 8, I put the Inferential Conception (IC) to work in the historical case study: the genesis of GR. I analyze three historical episodes, using the conceptual apparatus provided by the IC. In episode one, I investigate how the starting point of the application process, the "assumed structure", is chosen. Then I analyze two small application cycles that led to revisions of the initial assumed structure. In episode two, I examine how the application of "new" mathematics – the application of the Absolute Differential Calculus (ADC) to gravitational theory – meshes with the IC. In episode three, I take a closer look at two of Einstein's failed attempts to find a suitable differential operator for the field equations, and apply the conceptual tools provided by the IC so as to better understand why he erroneously rejected both the Ricci tensor and the November tensor in the Zurich Notebook.

# Contents

# Introduction

What is the role of mathematics in application to the world? Why is mathematics useful in solving empirical problems? The present thesis takes these questions as a starting point for analyzing some prominent historical cases of the application of mathematics. It is a contribution to the debate on the applicability of mathematics; it examines the interplay between mathematics and the world in three small case studies, and one big study.

The question as to why mathematics is applicable to the world has a long philosophical tradition – it goes back as far as Plato.[1] It is, therefore, all the more surprising that the debate on this problem is not nearly as extensive as debates on other classical questions in the philosophy of mathematics and science.

As some observers have noted, this is about to change; the last two decades have seen a renewed interest in the applicability of mathematics.[2] Very recently, detailed accounts of the interplay between mathematics and empirical problems have been proposed. What is more, the issue of applicability is at the core of some of the core debates in the philosophy of science; in particular the philosophy of physics.[3]

The focus of this thesis is not exclusively on the application of a (given) mathematical theory to a (new) domain of application, because I am also interested in what could be called *mathematization*, the creation of mathematics in view of one particular empirical problem, and, generally, the dynamics of the interaction between empirical problems and mathematics.

One of the guiding ideas of the applicability debate is that, at least part of, the usefulness of mathematics in application is due to the fact that we can use mathematical structures, or models, to represent relevant empiri-

[1]See Steiner (2005, p. 626) for a short discussion of historical examples.

[2]See e.g. Colyvan (2009) and Steiner (2005), two recent handbook articles on the issue, and the literature therein.

[3]To give an example, the question as to which aspects of the formalism of quantum mechanics should be interpreted realistically, is, at its core, the question as to how a particular mathematical representation is related to the world. A very similar problem crops up in the debate on the Hole Argument in general relativity – I discuss this case in chapters 5 and 8. It would be fruitful to bring the various strands of the debate together; at least in some cases, specialized debates could benefit from the general perspective provided by the applicability debate.

cal structure, and then exploit the inferential possibilities of mathematics, to gain knowledge about the world. The idea that structure-preservation is a relevant part of applicability is the main thesis of the so-called *Mapping Account*, while the emphasis on the inferential role of mathematics is at the core of the so-called *Inferential Conception* of the application of mathematics. Putting this simple account of application to work in real-life cases is one of the goals of this thesis.[4]

The idea that the applicability of mathematics is rooted in a structural correspondence between mathematics and the world raises many intricate issues. One of the basic questions concerns the nature of mathematical structures. What is a mathematical structure? When I began writing this thesis, I wanted to clarify this question before embarking on the more complex problem of how mathematical and empirical structures are related. Therefore, I first examined a promising structuralist proposal for the metaphysics of mathematics: Stewart Shapiro's so-called *ante rem* structuralism.

The upshot of this study is twofold. First, I am quite critical of *ante rem* structuralism, because it is, as I argue in chapter 1, in conflict with mathematical practice. More specifically, mathematicians ascribe properties to mathematical structures that are incompatible with *ante rem* structures. Secondly, I had to realize that the question as to how to characterize mathematical structures is itself worthy of a whole thesis, and that I therefore had to move on, if I also wanted to address the problem of applicability. The discussion of *ante rem* structuralism reinforced my belief that we have to pay close attention to mathematical and scientific practice. However, it did not lead to more than a sketch of how we might characterize mathematical structures.[5]

An important question in the discussion of applicability is how it meshes with classical positions in the metaphysics of mathematics. Does the fact that mathematics is applicable favor a Platonist, Nominalist, or a Formalist position? This question is the motivation behind the debates on the so-called *Indispensability Arguments* – I will sketch this debate very briefly below. I will not have much to say on the relation between applicability and the metaphysics of mathematics. My contribution to this debate is limited to a critical discussion of the examples used in this debate, and to some remarks

[4]See chapter 7 for a discussion of the Mapping Account and the Inferential Conception.

[5]The problem of what mathematical structures are, and the problem of applicability, are closely related. Here are two examples. First, there is a close connection between problems of characterizing purely mathematical structures, and problems that beset the genesis of GR. In both cases, some of the problems can be solved by accepting *isomorphism types* – in the case of GR, equivalence classes of metrics related by diffeomorphisms – as the prime representational tool of the "real" structure, i.e structure that is free of representational artifacts. Second, it has been proposed that we can ascertain applicability on the level of set theory, the "foundational theory" of mathematics, by admitting so-called *urelements*. I briefly argue, in chapter 1, that this can only account for a minority of applications.

on conceptual issues, which could be fruitfully applied in this context.[6]

I presuppose throughout the thesis that the domain of mathematics, and the empirical domain, are distinct, whatever their metaphysical status may be. This is compatible with Platonism, in that Platonists take mathematical entities to be abstract, i.e. neither spatiotemporally located, nor causally active, whereas the empirical domain is not abstract. It is also compatible with Nominalism, in that a Nominalist can maintain scientific realism, and distinguish the two domains on this basis.[7]

I think that the usefulness of mathematics in application is not a controversial thesis, but rather a real phenomenon that we should try to understand. However, it is in no way my goal to contribute to the mystification of mathematics – in particular in its application to physics – that has bewitched some philosophers, mathematicians and scientists.[8] Therefore, my interest is not limited to the successful application of mathematics. Quite to the contrary, I think that we have a lot to learn from failed application: cases where some mathematical approach, theory, or model is simply unsuitable for a particular empirical problem. Some of the cases I examine are examples of the *unsuccessful* application of mathematics.

## Scope and Method

It goes without saying that I cannot touch on all aspects of the applicability of mathematics; the systematic scope of the thesis, and the number of case studies, is very limited. Systematically, the focus is on the role of mathematics in theoretical models and model building (in chapter 2, the Königsberg case; in chapter 3, the predatory-prey model; and in chapter 4, the bee's honeycomb), and the genesis of a fundamental physical theory, in the case study on GR. I do not consider data-driven modeling, or the use of statistics, which are also cases of the application of mathematics. Furthermore, there is a focus on a potential explanatory role of mathematics in application. This is partially a function of the direction that recent philosophical discussions have taken. I also touch on other theoretical roles of mathematics in application, but these are not systematically evaluated.

Methodologically, the smaller case studies take the general question of applicability as a starting point. However, they are not primarily driven by particular philosophical questions and theses, but rather by the historical cases themselves. The idea is to get a firm basis of the underlying science in

[6] I critically discuss contributions proposing an explanatory role of mathematics to scientific explanations in chapter 2; chapters 3 and 4 also examine the role of mathematics in an explanatory context.

[7] I discuss the question as to whether the separation between the two domains is problematic in chapters 2 and 7.

[8] Here I am thinking of early contributions to the debate on the "Unreasonable Effectiveness of Mathematics"; I will briefly discuss this debate below.

the first step, and to extract insights about the interplay between mathematics and empirical questions directly from the cases. Only in the second step are the cases brought into contact with the recent philosophical discussion. The case study on GR is an exception, in that our reconstruction is based on a philosophical account of applicability.

The methodological approach of taking complex, historical cases as starting points has several advantages:

1. In previous philosophical studies, there was a tendency to work with simple case studies, which I call toy examples, which were, at times, not presented in sufficient detail. A more careful look at the case studies makes it possible to correct some misconceptions that have sneaked into the philosophical literature.

2. The in-depth analysis of the case studies makes it possible to develop novel systematic results. We are far away from having a clear picture of the variety of roles that mathematics can play in application, and we need to collect data. The case studies provide the raw material for the formation of novel philosophical hypotheses.

3. The detailed analysis of historical cases enables the study of the *process* of application. We can examine the genesis of a mathematically formulated empirical theory.

4. Some of the theses of the applicability debate have a historical component. In particular, Mark Steiner (1998) has proposed that at least a part of the "Unreasonable Effectiveness of Mathematics" lies in the effectiveness of mathematics in discovery, e.g., by suggesting the form of new theories, and by predicting new empirical phenomena. In order to address this issue, it is necessary to examine the historical genesis of scientific theories. We will do this in the case study on GR.

5. The study of historical sources reveals the messy details of actual science, as opposed to the smooth presentations of scientific models and theories in textbooks. These presentations tend to obfuscate the assumptions, idealizations, and compromises that are necessary for finding a mathematical formulation of a theory. However, this is exactly what we are interested in. This is what makes mathematics applicable in the first place.

## Applicability: Philosophical Debates

The thesis does not contain a survey of the literature on the applicability of mathematics.[9] It may therefore be helpful to sketch the relevant philosoph-

[9] I briefly review the relevant debate on mathematical structuralism in chapter 1.

ical debates, very briefly.

The debate on applicability has two main strands. The first strand is about the so-called "Unreasonable Effectiveness of Mathematics in the Natural Sciences", discussed in a famous paper by Eugene Wigner (1995). The second strand concerns the so-called "indispensability of mathematics" to empirical science, and goes back to remarks by Quine and Putnam. I will now, briefly, discuss both these strands.

The Unreasonable Effectiveness of Mathematics[10] is the thesis that the conjunction of the following two facts is surprising – or even miraculous – and in need of explanation: a) mathematics is invented, or discovered, based on aesthetic considerations, which are largely independent of empirical questions, whilst; b) mathematics is also the language in which many successful empirical, and, in particular, physical, theories are formulated. The puzzle is how is it possible that a tool that was developed independently of empirical considerations can be successfully applicable to real-world problems.

Wigner famously wrote that

> [t]he miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve. (Wigner, 1995, p. 549)

One way of accounting for this problem appears to be Platonism: if the world is inherently mathematical, then it is no wonder that mathematical discoveries lead to good empirical theories. However, as Mark Colyvan (2009, sec. 6.2) points out, this leaves many questions unanswered; for example why the method of mathematics, which is wildly different from empirical methods, is nevertheless helpful in the formulation of empirical theories. It is more obvious that Wigner's puzzle is a problem for formalist, or anti-realist, positions in the philosophy of mathematics.[11]

An important, sustained discussion of applicability is that provided by Mark Steiner (1998). Steiner points out that there is not just one problem of applicability, but rather that the puzzle can be given many forms, some of which he considers to be unproblematic. For example, he claims that the *semantical problem* of applicability, the question as to why arguments drawing on mathematical facts are valid, despite the fact that numbers seem to feature both as objects and as predicates, has been solved by Frege.[12]

Steiner considers other varieties of the problem to be more serious; for example the *descriptive problem*, i.e., the appropriateness of mathematical concepts in the description and solution of empirical problems. Steiner gives

[10] See Steiner (2005) and Colyvan (2009, sec. 6) for an overview; the following account draws on the latter.

[11] See Colyvan (2009) for further discussion.

[12] See Steiner (1998, p. 16) for an example and Frege's solution.

Wigner's puzzle a distinctively historical twist. He emphasizes that the role of mathematics in the discovery of empirical theories and phenomena is particularly puzzling: how can mathematics be helpful in the formulation of theories that lead to the discovery of new phenomena?

There are systematic problems with Wigner's and Steiner's puzzle. For one, it is simply not clear that mathematics really proceeds based on aesthetic considerations only. Also, if we wanted to gauge the real extent of the *unreasonable* effectiveness of mathematics, it is not sufficient to only take the successes of application into account; we would also have to consider the many failures. It is unclear how to carry out such a study, and whether the outcome would still speak in favor of a miracle.[13]

I will not systematically discuss Wigner's or Steiner's proposals in this thesis. The questions they raise are important, but I think it is not helpful to couch the problem in terms of an *unreasonable* effectiveness, or miracles. We all know that there are no miracles. I will, however, point out potential lines of conflict between their views and the cases I discuss; also, I will briefly return to the unreasonable effectiveness in the conclusion.

The second strand of the applicability debate is concerned with arguments for the indispensability of mathematics in application.[14] Very roughly, the original indispensability argument runs as follows[15]: We should accept exactly those entities as real that are indispensable for our best scientific theories. *Mathematical* entities are indispensable for the formulation of our best physical theories, and therefore, we should be realists about mathematical entities; i.e., we should be Platonists.

This argument is only of interest if we accept scientific realism; the argument has no traction if we simply deny the existence of the (unobservable) entities of science. The argument has the form of an inference to the best explanation (IBE) – the reality of mathematical entities explains why we need mathematics to formulate our best scientific theories.

One line of attack against the argument is to question whether all theoretical entities we use in the formulation of scientific theories are on a par, e.g. by postulating an additional criterion, which blocks accepting the reality of mathematical entities.[16]

A second line of attack was pursued by Hartry Field (1980).[17] Field accepts the IBE part of the argument, but denies that mathematical entities are indispensable for the formulation of our best scientific theories. In order to demonstrate the dispensability of mathematics, he nominalizes Newtonian gravitational theory by "geometrizing" the theory and thus eliminating quan-

[13]See e.g. Maddy (2007, sec. IV. 2. iii.) for a critical discussion of the puzzles.

[14]See Colyvan (2009, 2011) for introductions; Colyvan (2001) is a book-length discussion of indispensability.

[15]See Colyvan (2009, p. 656).

[16]See e.g. Maddy (1997).

[17]See Colyvan (2001, ch. 4) for an overview of Field's program, and criticism.

tification over real numbers, and thus demonstrates that everything derivable in a mathematical theory is also derivable in its nominalized counterpart.

Field's program is now commonly taken not to have been successful. One of the objections against Field's program is its lack of naturalism. His nominalized gravitational theory is simply not a real-life alternative to the classical formulation using mathematical entities, such as the potential function.

It is commendable that Field shifted the focus of the discussion to real-life examples, where mathematical entities are taken to be indispensable. One important example has been proposed by Alan Baker (2005): the explanation of the prime-numbered life-cycles of cicadas, using a theorem from number theory.[18] Baker, and the recent debate in general, has focused on one particular sense in which mathematics is indispensable, namely *explanatory* contribution of mathematics to science. Mathematics is taken to be indispensable to a scientific explanation, if the explanation using mathematics is superior to alternative explanations without mathematics. I will critically discuss examples from this debate in chapters 2, 3 and 4.

## Chapter Synopsis

Here is a short summary of the thesis.

### Part I

The first part of the thesis consists of four small case studies. The first is concerned with the metaphysics of mathematical structures, while the other three are about applicability.

- In chapter 1, I criticize *ante rem* structuralism, proposed by Stewart Shapiro, by showing that it is in conflict with mathematical practice. Shapiro introduced so-called "finite cardinal structures" to illustrate features of *ante rem* structuralism. I establish that, although these structures have a well-known counterpart in mathematics, this counterpart is incompatible with *ante rem* structuralism. I then discuss whether the controversial feature of *ante rem* structuralism is compatible with the most common representational tool of mathematical structures, set theory. Finally, I review the prospects for retaining features of *ante rem* structures by using isomorphism types.

- In chapter 2, I discuss Leonhard Euler's solution of the Königsberg bridges problem. I propose interpreting Euler's solution both as an explanation within mathematics and as a scientific explanation. The purely mathematical explanation is not a proof, but an application of Euler's theorem. I put this notion of intra-mathematical explanation

[18] See section 2.4.1 for a brief exposition and discussion.

to work against two recent philosophical accounts of the Königsberg case; Alan Baker (2012) and Marc Lange (2013).

I then discuss whether the applied version of Euler's solution can be interpreted as a causal explanation. I suggest that, on a broad reading of the notion, this is a causal explanation. Finally, I claim that a pragmatist reading of the relation between the mathematical formulation of the problem, and its real-world counterpart, based on Bas van Fraassen's theory of explanations, allows for an adequate understanding of the case. I examine two philosophical analyses of the Königsberg case, in view of the preceding analysis. I argue that the proposal by Christopher Pincock (2007), to reconstruct the case as an "abstract explanation", is incomplete, and that the account of "distinctively mathematical explanations" by Marc Lange (2013), which goes against a causal reading of the case, conflates the applied and the purely mathematical versions of Euler's solution.

- In chapter 3, I analyze the predator-prey model, proposed by Lotka and Volterra in the beginning of the 20th century. I revisit the historical papers by Vito Volterra. After following the historical discussion of the model in time, and giving a brief account of the model's status in population ecology today, I extract some interesting philosophical lessons from Volterra's original account of the model, such as: Volterra's remarks on mathematical methodology; the relation between mathematics and idealization in the construction of the model; some relevant details in the derivation of the Third Law, and; notions of intervention that are motivated by one of Volterra's main mathematical tools, phase spaces.

  I put the analysis to work in some recent philosophical debates on the predator-prey model. I argue that Mark Colyvan (2013) underestimates the importance of idealizations in the model. I then critically examine the recent contribution by Weisberg and Reisman (2008) to the debate on robustness analysis. I claim that Weisberg and Reisman's account suffers from mathematical imprecisions. I also cast doubt on robustness analysis as a phenomenon unique to biology, as opposed to other modeling sciences. Finally, I argue that the analysis by Christopher Pincock (2012), according to which the predator-prey model is an acausal representation, is mistaken.

- In chapter 4, I discuss scientific and mathematical attempts to explain the structure of the bee's honeycomb. The chapter has two parts. In the first part, I discuss a candidate explanation, based on the mathematical Honeycomb Conjecture, presented in Lyon and Colyvan (2008). I argue that this explanation is not scientifically adequate for two reasons. Firstly, I show that the explanation is deficient because the HC solves a two-dimensional problem, whereas an actual honeycomb has a three-dimensional structure that cannot be adequately captured in two dimensions. I then cast doubt on the idea that we should accept the HC even as a partial explanation of the actual, three-dimensional honeycomb.

In the second part, I discuss other mathematical, physical and biological studies that could contribute to an explanation of the bee's honeycomb. I examine a mathematical explanation proposed by Laszlo Fejes Tóth (1964). I argue that the mathematical result of this account is not applicable to the bee's honeycomb because one of the idealizations it introduces is too strong. I review some recent biological investigations of the bee's honeycomb, including Pirk et al. (2004); Hepburn et al. (2007); Bauer and Bienefeld (2013). I then call attention to an alternative explanation of the bee's honeycomb based on these results. Finally, I introduce a general framework that classifies the bee's honeycomb as a kind of foam, and I give a short account of an experiment by Weaire and Phelan (1994), that can be interpreted as a physical realization of the bee's honeycomb. The upshot is that most of the relevant mathematics is not yet sufficiently understood, and there is also an ongoing debate as to the biological details of the construction of the bee's honeycomb. However, the results from the physics of foams, depending on the outcome of the biological debate, could provide an explanation.

## Part II

The second part of the thesis is a bigger case study from physics: the genesis of GR. Chapter 5 introduces the necessary historical and systematics background from physics and mathematics; chapter 6 is a historical study of the "new" mathematics applied in GR, and Marcel Grossmann's contribution to its application, and; chapter 7 introduces, discusses and extends the Inferential Conception, the account of the application of mathematics that we use in the analysis of the case. Chapter 8, finally, is where rubber – the IC – meets the road – GR.

Some brief readings instructions are in order. Most importantly, it is not necessary to read all of chapter 6 for the philosophically-minded reader; it is sufficient to read the highlights in section 6.10. For those familiar with the history of GR, a short glance at chapter 5 should be sufficient.

Also, it should be noted that this part of the thesis, is based on joint work with Tilman Sauer, except for the introductory chapter 5. This means that, while I am the author of these three chapters, they are based on our discussions, and have been substantially improved by extensive comments by Tilman. This is why I have switched to the plural form when referring to the authors.

- Chapter 5 is a short introduction to the history, physics and mathematics that is relevant to the genesis of GR. I first introduce physical and mathematical theories that were available before the search for GR started, and then I sketch the tumultuous genesis of GR in the years 1907 – 1916, with a focus on the transition to the mathematical theory, which is now thought to be indispensable for the formulation of GR, tensor calculus, and the adaptation of that theory to gravitational theory.

- Chapter 6 discusses the historical question as to what Marcel Grossmann contributed to the genesis of GR. We will examine the so-called "Entwurf" paper, an important joint publication by Einstein and Grossmann, containing the first tensorial formulation of GR. In particular, we will analyze the second, mathematical part of the Entwurf, and we will discuss the origin of the mathematical theories used in this part, as well as Grossmann's own, novel contributions.

  The Entwurf theory constitutes the earliest meeting point of the historical predecessor of tensor calculus, the "Absolute Differential Calculus" (ADC), and gravitational theory. Previously, the ADC had been developed independently from any application to gravitational theory. By comparing Grossmann's part with the mathematical theories he used, we can gain a better understanding of what is involved in the first steps of assimilating a mathematical theory to a physical question.

- In chapter 7, we introduce and discuss a recent account of the applicability of mathematics to the world, the Inferential Conception (IC), proposed in Bueno and Colyvan (2011). The chapter has three objectives. First, we give a short exposition of the IC, which improves on a previous account of applicability, i.e. the mapping account. Then, we offer some critical remarks on the account, and discuss potential philosophical objections. Third, we propose some extensions of the IC, preparing the ground for the application of the IC to our case study in chapter 8.

- In chapter 8, we put the Inferential Conception (IC) to work in our historical case study: the genesis of GR. We analyze three historical episodes, using the conceptual apparatus provided by the IC. This, in turn, will help us refine the account.

  In episode one, we investigate how the starting point of the application process, the "assumed structure", is chosen. We will clarify the status of the starting point of the application process, and discuss the trigger of the application process. Then we analyze two small application cycles that led to revisions of the initial assumed structure.

> In episode two, we examine how the application of "new" mathematics – the application of the Absolute Differential Calculus (ADC) to gravitational theory – meshes with the IC. We describe how the mathematical part of the Entwurf is shaped by the application process. Our focus is on the application cycle that led to the "discovery", and the application, of the ADC, i.e. the quest for generally covariant differential operators.
>
> In episode three, we will take a closer look at two of Einstein's failed attempts to find a suitable differential operator for the field equations, and apply the conceptual tools provided by the IC so as to better understand why he erroneously rejected both the Ricci tensor and the November tensor in the Zurich Notebook.

I end with a conclusion, in which I comment on some of the common themes of the case studies.

# Acknowledgements

I am very grateful to Michael Esfeld for accepting me as his student, and for his continued support. One of my goals was to reach his standard of philosophical clarity, which, I hope, is demonstrated in the present work. I am also grateful to the members of the jury, Hannes Leitgeb and Claus Beisbart, for reading and commenting on my thesis.

I have profited immensely from the collaboration with Raphael Scholl. The long discussions I had with Raphael during the last few years have profoundly shaped my perspective on virtually all philosophical topics. I would not have enjoyed writing the present thesis nearly as much, if it had not been for him.

I am very much indebted to Tilman Sauer, who helped me understand the intricacies of the history of GR. His intellectual generosity and openness do not cease to amaze me; I could not have written the second part of the present thesis without him.

There are various people who have read papers or chapters that ended up as parts of the present thesis, or who have helped me with discussions. Thanks to Mark Colyvan, Matthias Egg, Martin Gasser, Marion Haemmerli, Michael Messerli, Philip Mills, Antoine Muller, Thomas Müller, Kärin Nickelsen, Christopher Pincock, Christian Sachse, Ilona Stutz, Adrian Wüthrich, and various anonymous referees. Special thanks to Dan Ward for proofreading.

Finally, I would like to thank my friends for their moral support, in particular during the final stage of writing up the present thesis.

This work was supported by the Swiss National Science Foundation, grant numbers (100011-124462/1) and (100018-140201/1).

# Part I

# Small Case Studies

# Chapter 1

# Against Ante Rem Structuralism

## 1.1 Introduction

When it comes to the nature of mathematical objects, many philosophers and mathematicians embrace a form of structuralism. Philosophers of mathematics have tried to formulate a metaphysics of mathematics in structuralist terms for quite some time. One kind of structuralism is particularly popular: Stewart Shapiro's *ante rem* structuralism, first proposed in Shapiro (1997).

In this chapter, I critically assess *ante rem* structuralism. After a short introduction to *ante rem* structuralism in section 1.2, I raise my principal objection to this position in section 1.3 by showing that it is in conflict with mathematical practice. Shapiro introduced so-called "finite cardinal structures" to illustrate features of *ante rem* structuralism. I establish that these structures have a well-known counterpart in group theory, but this counterpart is incompatible with *ante rem* structuralism: It has an *in re* character. Furthermore, there is a good reason why, according to mathematical practice, these structures do not behave as conceived by *ante rem* structuralism: We want to be able to establish connections between different representations of abstract structures, and in order to do this, we rely on "coordinates", non-structural properties of structures.

In section 1.4, I discuss the role of set theory for the *in re* perspective on structures. It seems to me that the fact that domains of structures are commonly taken to be sets can be explained by the fact that sets naturally provide "surplus structure" and thus serve as coordinates of structures. This is so because, on the most common conception of sets, the elements of sets are discernible. I go through several notions of set theory to show that this is the case.

The set-theoretic construction of sets is not completely satisfactory. Set-theoretic representations of structures have properties that the structures do not have intrinsically – "surplus structure". However, there are ways to deal with this problem, which I discuss in section 1.5. The solution is based on the notion of isomorphism types, which allows us to separate those properties that are due to a particular representation from those properties that belong to the represented structure. This solution will not satisfy the *ante rem* structuralist, as it cannot do away with all the "surplus structure", but it is probably as close as we can go towards *ante rem* structuralism: We get *ante rem* properties and relations, but not *ante rem* structure.

## 1.2 Ante Rem Structuralism

In this section, I first give a very short introduction to *ante rem* structuralism. Then I explain a relevant objection that has been raised against an early version of *ante rem* structuralism – that it endorses a version of the principle of the identity of indiscernibles – and show how Shapiro was able to avert this problem. Finally, I lay out a feature of *ante rem* structuralism that I find to be troubling, the fact that we cannot name places of certain symmetric structures.

### 1.2.1 The Idea

One starting point of mathematical structuralism is Benacerraf (1965). Benacerraf raises challenges to ontological realism in mathematics. First, he argues that numbers cannot be particular sets, because we have no sufficient reason to identify the natural numbers with one particular set-theoretic representation instead of another, e.g. Zermelo or von Neumann ordinals.[1] Second, this argument is extended to objects in general - numbers cannot be objects. What, then, do we talk about when we talk about natural numbers? Besides the negative answers just sketched, Benacerraf hints at a positive, structuralist answer:

> [...] in giving the properties [...] of numbers, you merely characterize an *abstract structure* - and the distinction lies in the fact that the 'elements' of the structure have no properties other than those relating them to other 'elements' of the same structure. (Benacerraf, 1965, p. 285)

This idea has been taken up by structuralists, most notably in Shapiro (1997). According to his *ante rem* structuralism, we should not think of mathematical structures in terms of their instantiations (*in re*), but in terms

[1] The von Neumann ordinals start with $\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, ...$, the Zermelo ordinals with $\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, ...$

of the structural features that the structures have independently of, or "before", instantiations (*ante rem*). For example, *the* structure of natural numbers is independent of its instantiations as, say, some ordinal structure. It is exhaustively characterized by the axioms of natural numbers. The natural numbers are the places in this structure, characterized in terms of the structural relations, such as the successor function. This is the so-called places-are-objects perspective of *ante rem* structuralism.

*Ante rem* structuralism is an attractive position, because each mathematical structure is taken seriously in itself. We work exclusively with the properties and relations that are naturally available in a structure; it is not necessary to interpret structures in terms of, say, set theory. This meshes well with many mathematicians's conception of the autonomy of mathematical subdisciplines: A graph theorist is working with graphs, he is not doing some version of set theory.

### 1.2.2 An Objection: PII

A serious objection, however, has been raised against *ante rem* structuralism; see Burgess (1999) and Keränen (2001). The objection is based on two facts. According to *ante rem* structuralism, we can characterize mathematical objects exclusively in terms of the structural properties (including relations) of the structure to which the objects belong. Secondly, Shapiro can be read as endorsing a form of the Principle of the Identity of Indiscernibles (PII): if two objects of a structure share all structural properties, then they should be identified; see Shapiro (2008, p. 286). This leads to the objection that structures with certain symmetries are not adequately captured by *ante rem* structuralism.

The concept of structures exhibiting more or less symmetry can be made more precise using the concept of *automorphism*. An automorphism is a structure-preserving function (isomorphism) from the structure to itself. In the case of natural numbers, there is only one automorphism, the identity function. Structures on which only this (trivial) automorphism can be defined are called *rigid*. Structures admitting of non-trivial automorphisms are called *non-rigid*. Places of structures linked by a non-trivial automorphism are called *structurally indiscernible*.

Non-rigid structures, such as the complex numbers, do have places, e.g. $i$ and $-i$, that are structurally indiscernible – but which are nonetheless not identical: the additive inverse of $i$ is $-i$, not $i$ itself. As 0 is the only complex number additively inverting itself, and $i$ is not 0, $i$ and $-i$ have to be different. But, according to (IND), $i$ and $-i$ should be identified.[2] *Ante*

[2]Note that $i$ and $-i$ need only be identified according to (IND), which is just one possible formulation of PII. There are several notions of indiscernibility on the market; see Ketland (2011) and Ladyman et al. (2010) for a discussion of these notions and their interrelations). According to *weak discernibility*, $i$ and $-i$ are discernible by a formula

*rem* structuralism appears not to adequately capture mathematics, which is unacceptable for a nonrevisionist position, such as Shapiro's.

In reaction to this objection, Shapiro (2008) agrees that it would be fatal if *ante rem* structuralism were committed to the above form of PII. However, he denies that this is the case. He thinks that in mathematics, identity cannot be defined in a non-circular way, and that mathematics presupposes identity.[3] *Ante rem* structuralism can thus be amended in the following way: we use only structural properties that are naturally available in a mathematical structure to characterize the objects belonging to that structure, and identity is one of these structural properties.

If we accept identity as a primitive relation, then Shapiro has successfully averted attacks based on PII. For the sake of the argument, I accept Shapiro's solution and assume that identity is available as a primitive relation. What follows has nothing to do with metaphysically motivated principles, such as PII.

### 1.2.3 No-Name Places

The feature of *ante rem* structuralism that I consider to be problematic concerns reference in mathematics. To see the problem more clearly, I will underline an implicit distinction made by Shapiro.

In Shapiro's opinion, one attractive feature of *ante rem* structuralism is that "in most cases, reference is straightforward" (Shapiro, 2008, p. 290). One straightforward case is the structure of natural numbers with unique, structurally characterized places interpreted as objects: the numeral "4" refers to the fifth place in this structure. While Shapiro accepts the idea that "singular terms in true sentences [... suggest] that there are objects denoted by those terms", he denies the converse: "It is simply false that to be an object is to be the sort of thing that can be picked out uniquely with a singular term." (Ibid.) How can this be the case?

Shapiro gives several examples where reference to mathematical objects fails. For big structures, such as the real numbers, at least one problem of reference is well-known: Given a countable supply of names, we cannot name or describe all real numbers at once, as they are uncountable. We can "diagonalize out" of any list of members of these structures. Therefore, the countable supply of names cannot be in a one-one-correspondence with the members of these structures. There are probably further problems with reference to members of big or random structures, but I will not discuss them further, as the claim about failure of reference due to uncountability

$\phi(x, y)$ expressing the fact that $x$ and $y$ are additive inverses: $\phi(i, i)$ is false in the complex number structure, while $\phi(i, -i)$ is true. If (IND) were based on weak discernibility, then $i$ and $-i$ would be discernible (but see Ketland (2006) for criticism of weak discernibiliy).

[3] See Shapiro (2008, p. 292). The proposal that identity is presupposed in mathematical practice has been made in Ketland (2006), as Shapiro notes.

is uncontroversial.

My focus will be on a different type of example: structures that are *too homogeneous* for reference. Shapiro thinks that certain mathematical structures with symmetries have the property that we cannot name or refer to the objects, or places, in these structures because they are *too homogeneous*. He writes:

> There simply is no naming *any* point in Euclidean space, nor any place in a finite cardinal structure and in some graph, no matter how much we idealize on our abilities to pick things out. The objects are too homogeneous for there to be a mechanism, even in principle, for singling out one such place, as required for reference, as that relation is usually understood.(Shapiro, 2008, p. 291)

I take it that the reason why we cannot name the objects in these structures is that there are no structural properties to pick them out, or discern them. Identity is of no help, as structurally indiscernible places can be nonidentical. I will call this the "no-naming constraint" of *ante rem* structuralism.

I think that the no-naming constraint is an undesirable feature of *ante rem* structuralism, because it is in conflict with mathematical practice. This I will show by examining the paradigm of homogeneous structures, the finite cardinal structures mentioned in the above quote. Finite cardinal structures comply with the no-naming constraint to the extreme: none of their places can be named, because they are too homogeneous. I will show that the correlate of finite cardinal structures in mathematics does not comply with the no-naming constraint.

This creates a problem for Shapiro, because he also endorses the so-called *faithfulness constraint*. This is the "*desideratum* [...] to provide an interpretation that takes as much as possible of what mathematicians say about their subject as literally true, understood at or near face value" (Shapiro, 2008, p. 289, emphasis in original).[4] Shapiro wants his position to be in agreement with mathematical practice as much as possible. If naming the objects of finite cardinal structures is no problem in practice, then either the no-naming constraint or the faithfulness constraint has got to go.

## 1.3 An Objection from Group Theory

### 1.3.1 Finite Cardinal Structures in Mathematics

Shapiro characterizes the cardinal-four structure, one kind of finite cardinal structure, as follows:

[4] See Shapiro (1997, ch. 1) for more on the faithfulness constraint.

> The cardinal-four structure [...] has four places, and *no relations*. [...] Since there are no relations to preserve, every bijection of the domain is an automorphism. Each of the four places is structurally indiscernible from the others and yet, by definition, there are four such places, and so not just one. (Shapiro, 2008, p. 287)

The cardinal-four structure has four objects, or places, and no relation between these objects; therefore, every bijection between the objects is an automorphism. Technically speaking, the places are pairwise structurally indiscernible.

We will now locate the cardinal-four structure in mathematical practice. Initially, it is unclear how to interpret the cardinal-four structure in ordinary mathematical terms, because if we cannot name the objects of a structure, it is not clear how to define a function on the structure.[5] We will therefore choose a familiar starting point, and work our way from there. We will use the familiar idea that structures can be characterized via structure-preserving functions.

Usually, a structure is defined by giving some domain, say $C = \{1, 2, 3, 4\}$, on which we can define functions in the usual way. There are no relations on $C$, so every $f : C \to C$, with $f$ bijective, is an automorphism. In mathematics, a bijection on a (finite) domain, which is not required to respect any relations, is called a permutation of $C$. Mathematicians are interested in permutations because the set of permutations of a (finite) domain, equipped with composition of functions, forms an important group called *symmetric group*, written $S_n$ if the size of $C$ is $n$. The members of the group $S_4$ are the permutations of $C$.[6]

Clearly, $C$ is not the cardinal-four structure: the elements of $C$ are natural numbers, thus we can name them. This carries over to the permutation group on $C$: According to the *ante rem* structuralist, some of the permutations of $C$ should be indistinguishable. Take the functions $f$, defined as $f(1) = 2, f(2) = 1, f(3) = 3, f(4) = 4$, and $g$, defined as $g(1) = 3, g(2) = 2, g(3) = 1, g(4) = 4$. They are different members of $S_4$. However, the only difference between $f$ and $g$ is that $f$ permutes 1 and 2, while $g$ permutes 1 and 3.

Thus we cannot use the permutation group $S_4$ to characterize the cardinal-four structure: $f$ and $g$ are distinguishable, which should not be the case, as 2 and 3 play the same structural role. We have to "identify" $f$, $g$, and any other permutation of $C$ that only swaps two places of $C$ and leaves all other places untouched.

[5]This concern has been formulated before, see Hellman (2005, p. 545, fn. 10).

[6]The portion of elementary group theory used in the following can be found in any introduction to group theory, see e.g. Rotman (1995).

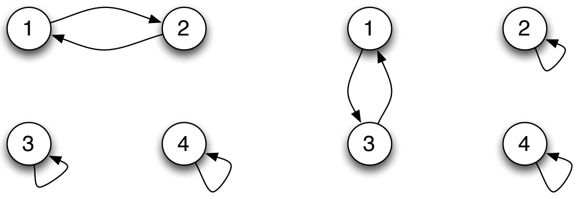



Figure 1.1: Functions $f$ (left) and $g$ (right)

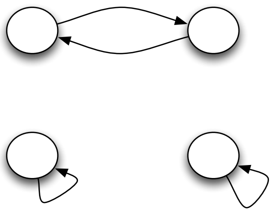


Figure 1.2: Function on Cardinal-Four Structure

In mathematics, the result of this identification is well-known and has many different, but equivalent descriptions. A particularly intuitive approach is based on the notion of cycles. To understand how this works, we need the cycle notation of permutations. A cycle of length $r \leq n$, written $(i_1 i_2 ... i_r)$, is a permutation $a \in S_n$ such that $a(i_1) = i_2, a(i_2) = i_3, ..., a(i_{r-1}) = i_r, a(i_r) = i_1, a(i_k) = i_k$ for $k \neq 1...r$, i.e. it sends $r$ places of the domain around in a cycle and leaves the other places alone. For example, $f$ above is the cycle (12). Alternatively, we can also write $f$ as (12)(3)(4), i.e. 3 and 4 are cycles of length one.

It is a theorem of group theory that all permutations can be written as a product of disjoint cycles. Thus we can think of permutations as cycles. This is very useful for our purposes, because we can use the cycle notation to classify all permutations into *cycle types*, also known as cycle structure. The cycle type of a permutation only depends on the number of cycles of length one, two, etc. of the permutation. The cycle type of a permutation $a \in S_n$ is written $(1^{m_1}, 2^{m_2}, ..., n^{m_n})$, meaning that the permutation $a$ has $m_1$ cycles of length 1, $m_2$ permutations of length 2, and so on. For example, the permutation $(12)(3)(4) \in S_4$ above is of type $(1^2, 2^1, 3^0, 4^0)$. Some thought reveals that the permutations in $S_4$ fall into five cycle types. Here is one instance of each type: (1)(2)(3)(4), (12)(3)(4), (123)(4), (1234), (12)(34).

I suggest that cycle types capture finite cardinal structures. A cardinal-four structure is completely characterized by the fact that we can define five "essentially different bijections" on its places: all permutations are automorphisms, but some permutations have to be identified. Now, the five essentially different bijections on the cardinal-four structure coincide with the five cycle types of $S_4$, see figure 1.3 below. These cycle types capture kinds of permutations by abstracting from the particular numbers (or places) that are permuted. They only appeal to facts, such as the number of places mapped to themselves, the number of places mapped to each other, the number of places mapped in three-cycles, and so on – the structure of cycles.

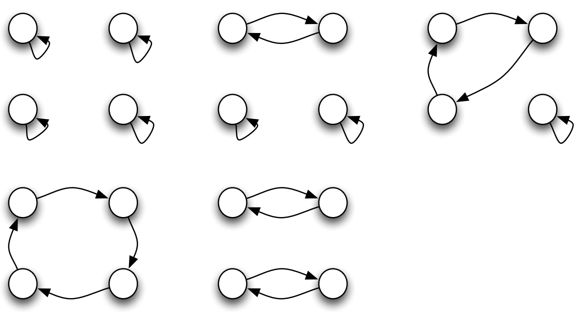


Figure 1.3: Cycle Types of $S_4$

It is not necessary to capture finite cardinal structures in terms of cycle types; the idea can be restated in many different forms. For example, there is a natural correspondence between cycle types and certain subgroups of $S_n$ called conjugacy classes: two permutations are of the same cycle type if and only if they are in the same conjugacy class.[7] Another perspective is in terms of partitions of natural numbers. A partition of $n$ is a sequence of natural numbers $i_1, i_2, ...i_r$ such that $i_1 + i_2 + ... + i_r = n$.[8] There is a one-one correspondence between partitions of $n$, cycle types of length $n$, conjugacy classes of $S_n$, and "essentially different" bijections on $C$, and we can use any one of these concepts to capture finite cardinal structures.[9]

### 1.3.2 Ante Rem Structuralism vs. Mathematical Practice

After this detour into mathematics, we are ready for our philosophical problem. Are cycle types *ante rem* structures in Shapiro's sense or not? More specifically, is it possible to name their places? Mathematical textbooks do not give a direct answer to this question, because naming is not a mathematical notion. However, they give an indirect answer.

[7]If $G$ is a group and $a$ a member of $G$, the conjugacy class of $a$ is the set of $b$ such that $b = xax^{-1}$ for some $x$ in $G$.

[8]Note, incidentally, that a closed form expression for $p(n)$, the number of partitions of $n$, is not known; see Simon (1996, p. 96). By extension, the same is true for the number of "essentially different" bijections on finite cardinal structures.

[9]Shapiro (2008) points out a suggestion in Leitgeb and Ladyman (2008) according to which finite cardinal structures are (isomorphic to) certain graphs. This is yet another way to conceive of finite cardinal structures.

In mathematics, different permutations such as $f$ and $g$ that belong to the same cycle type are always distinguishable. This follows from the way in which the permutations belonging to a certain cycle type are counted. A theorem tells us that the number of permutations of type $(1^{m_1}, 2^{m_2}, ..., n^{m_n})$ is $n!/\big(\Pi_{j=1}^{n}(m_j)!j^{m_j}\big)$; see Simon (1996, Theorem VI. 1.2.). Applied to the cycle type $(1^2, 2^1, 3^0, 4^0)$ of $f$ and $g$, we find that the number of permutations of this type is $4!/\big(2!1^2 \cdot 1!2^1 \cdot 0!3^0 \cdot 0!4^0\big) = 6$; these are the 6 permutations of the set $C$ that swap two places and leave two places untouched. This means that we can recover all the cycles belonging to a cycle type, all the different bijections on $C$, and especially $f$ and $g$. We can move freely between cycle types, as in figure 1.2, and cycles, as in figure 1.1.

This is not so according to *ante rem* structuralism. In figure 1.2, the places 2 and 3, while nonidentical, are structurally indiscernible. There are no properties or relations to discern them, and we cannot name them. From the perspective of *ante rem* structuralism, there is exactly one function in figure 1.2. But there is no way to recover, or count, different permutations, such as $f$ and $g$, in figure 1.1 that instantiate the function in figure 1.2.

The reason for this is that the *ante rem* structuralist can only use structural differences and identity to distinguish between $f$ and $g$. However, they have the same structural role: they swap two places, and leave two places alone. In particular, the *ante rem* structuralist cannot use the fact that $f$ and $g$ are different, because $f$ swaps 1 and 2, while $g$ swaps 1 and 3. All that can possibly matter for the *ante rem* structuralist is that two (nonidentical) places are swapped, while two further places, not identical to the former two, are left alone. There is one such situation, not two, or six.

Now, the *ante rem* structuralist could maintain that it is a primitive fact of identity that $f$ and $g$ are different permutations. However, this is not a fact that can be grounded in the identity and structural discernibility of the *places* that are permuted. Both $f$ and $g$ swap two nonidentical, structurally indiscernible places, so the nonidentity of places is of no help in distinguishing the two. The *ante rem* structuralist would need *additional* facts about the identity and discernibility of *functions*; more specifically, he would have to assume that there are exactly six different permutations with the same cycle type as $f$ – and so for all other cycle types of all permutation groups. I think this is not an attractive option.

The easy way out would be to state the obvious: $f$ and $g$ are different because, well, 2 and 3 are different. This, however, the *ante rem* structuralist cannot do, as he would have to label the places of the cardinal-four structure as 1, 2, 3, 4, and then describe the different permutations between these numbers, which, arguably, amounts to naming the places. Once he has adopted the *ante rem* perspective, the *ante rem* structuralist cannot move freely from figure 1.2 to figure 1.1.

Why is the situation different for the mathematician? The mathematician simply uses non-structural properties to discern the places of cycle types; for example, by defining the permutations on a set of natural numbers. He can then use the non-structural properties of the places of permutations to calculate how many permutations belong to each cycle type.[10] It appears that mathematicians adopt an *in re* perspective for circle types: they are not considered in isolation from their instantiations, but in close correspondence.

It could be asked[11] whether the problem of distinguishing the functions $f$ and $g$ could be solved by treating the places of the cardinal-four structure as parameters. We could skolemize the axiom of the cardinal-four structure

$$\exists x_1, x_2, x_3, x_4(x_1 \neq x_2 \wedge \ldots \wedge x_3 \neq x_4 \wedge \forall y(y = x_1 \vee y = x_2 \vee y = x_3 \vee y = x_4))$$

by eliminating the outermost existential quantifiers by introducing a new parameter for each quantifier. This procedure is akin to the rule of existential instantiation. If we now conceive of the functions as being defined on these parameters, we can very well distinguish the functions $f$ and $g$.

I think that this does not solve the problem, for the following reason. I certainly agree that the parameters can be used to *represent* the places of the cardinal-four structure. However, we can only distinguish the functions $f$ and $g$ as functions between parameters, which represent the places. If we want to establish, additionally, that we the functions $f$ and $g$ between the *places* are distinguishable as well, we would need a stable relation, a one-one correspondence between parameters and places (an interpretation of the parameters). This, however, would essentially amount to naming the places using parameters, which is impossible according to *ante rem* structuralism.

### 1.3.3 Ante Rem vs. In Re

My objection against *ante rem* structuralism is not *a priori* or metaphysical. The problem is Shapiro's faithfulness constraint, which it is in tension with the no-naming constraint; I argued that the no-naming constraint has consequences that contradict mathematical practice. Of course, the *ante rem* structuralist can claim that cycle types and the other structures above do not really capture his idea of finite cardinal structures. However, these structures are as close as mathematics gets to Shapiro's finite cardinal structures. If he does not think that cycle types adequately capture his idea, we can reasonably question the relevance of these structures for mathematics – unless he comes up with a mathematical structure that captures finite cardinal structures even better.

[10] The point that, in mathematics, we use non-structural properties to discern places in structures has been made before; see e.g. Hellman (2001). Hellman's criticism of *ante rem* structuralism is more general and severe than the one advanced here, as he considers the position to be incoherent.

[11] I thank an anonymous referee for this question.

According to Shapiro, the faithfulness constraint is relative and has to be weighted against other desiderata. If the feature of mathematics that is not faithfully mirrored by *ante rem* structuralism is only of minor importance, we could still dismiss it; after all, *ante rem* structuralism is able to capture some aspects of mathematical practice. Are there good reasons for conceiving of structures as *in re* rather than *ante rem*? Why is it important to count cycles of a certain type in a certain way?

There are good reasons for adopting an *in re* perspective. One reason is that it is an important part of mathematics to explore different perspectives, or representations, of one and the same abstract, *ante rem* structure. We saw an example of this practice above: we can think of finite cardinal structures in terms of cycle types, but also in terms of conjugacy classes or partitions of natural numbers. One advantage of these different representations is that we can use our knowledge of one of the representations for all the others.

However, in order to do this, we have to be able to prove that the different representations are equivalent, and in these proof, we often use instances of abstract structures ("Let $\pi$ be a permutation of type $x$ ..."), and structure-preserving mappings between these instances. This is why it is important that we can move freely between an abstract structure and its instances. This is impossible if we adopt an *ante rem* perspective, as we saw in the case of the cardinal-four structure.

Summing up, *ante rem* structuralism is right in emphasizing that we should take abstract mathematical structures seriously – they are more than their instantiations. However, we should not take abstraction too far. If we start to think of abstract mathematical structures as completely freestanding and independent of their instantiations, we lose sight of the fact that mathematics is also about the different representations of structures. If we want to make use of these representations, we have to be able to move back and forth between abstract structures and their instantiations, i.e. between an *ante rem* and an *in re* perspective.

## 1.4 Set Theory and Discernibility

In the previous section, I argued that mathematicians seem to adopt an *in re* perspective on structures, and that one reason why they adopt this perspective is that it makes it possible to establish relations between different structures. The fact that mathematicians want to be able to relate different structures goes a long way towards explaining why there is one particular, pervasive conception of structures in mathematics, namely a set-theoretic conception.

We usually define structures by first specifying a set, called the domain of the structure; on this domain, we can then define relations, functions, constants, and so on. Thus, on the one hand, the domain provides us with the substrate to define all these structural features. On the other hand, domains help us to establish connections between different structures by establishing mappings between them – it serves as a sort of coordinate system of the structure.

It is well-known that set theory can help clarify the notions of relation, function, and so on. I suggest that the second role of the domain, its usefulness in establishing connections between different structures, is also tied to the fact that the domain is a set. Sets have just the properties needed to serve as coordinate systems. The property that is at the core of the conception of sets is arguably extensionality: A set is constituted by (nothing but) the (distinct) elements of that set. We do not require any specific relations etc. to hold between these elements, except that they are distinct. Thus, in some sense, a set is as structureless as it gets: its structure is constituted by the number of its elements, i.e. its cardinality.

However, the distinctness of the elements of a set would not be sufficient to play the role of coordinate systems. For this, we need to be able to name them, and the discernibility of the elements of a set is a feature on which we rely to do so. In order to see whether the elements of a set are in fact discernible, we will scrutinize some conceptions of a set, gauge their overall relevance, and answer the question whether the elements of a set are discernible, and if yes, in what sense. As we will see, there is no uniform answer to this question; however, we can discern a tendency.

### 1.4.1 Sets in Mathematical Practice

We start with sets in mathematical practice, as this is the conception of sets that Shapiro would consider to be relevant. As we saw above in section 1.3, mathematicians will use domains such as $\{1, 2, 3, 4\}$ or $\{a, b, c, d\}$; obviously, natural numbers and letters are not only distinct, but discernible. Natural numbers are even structurally discernible in the structure of natural numbers. The discernibility of the elements of the two sets is useful for the reasons pointed out above: it makes it easy to define functions on these domains, and it makes it easy to define functions between the domains.

It could be objected that we cannot settle the question in this way. Maybe, mathematicians are, in some cases at least, sloppy in that they use sets with whatever elements they like in their daily work, be it numbers, letters, or anything else, and it could be a mistake to attach too much significance to this fact – the elements of these sets are discernible, but this need not be so in general. It could be instructive to have a look at more formal conceptions of set theory, in the spirit of an explication of the concept of sets, to find out whether the discernibility of the elements of a set is part of the core concept of sets or not.

### 1.4.2 Cantorian Set Theory

The first historical definition of sets was proposed in Cantor (1895, p. 481, my translation):

> By a 'set' we mean any collection $M$ of definite, well-distinguished [wohlunterschieden] objects $m$ of our intuition or our thought (which are called the 'elements of $M$) into a whole.

The important thing to note here is that the objects, or elements, are characterized as "wohlunterschieden". This does not just mean that they are distinct, but that they can be distinguished from each other, as emphasized by the prefix "wohl". It seems, then, that on the original conception of sets, the elements of a set are discernible. However, it is not yet clear how elements of a set can be discerned; the discernibility is merely postulated. This is different on at least some axiomatic set theories.

### 1.4.3 Pure Set Theory (ZF)

In most common, contemporary axiomatic set theories, such as ZF, the scope of what is accepted as elements of a set is very limited: sets and sets alone are sufficient to do set theory. It is not necessary to accept any other kind of object as an element of a set. This is an important difference from Cantor's formulation. Set theories in which this restriction is adopted are called pure set theories, as opposed to set theories with urelements, or atoms: these are mathematical or physical objects that are members of a set, while not being sets themselves.

In standard pure set theories, such as ZF, all sets are *structurally discernible qua sets*. Intuitively, the reason for this is that if two pure sets are nonidentical, then, by extensionality, there is a set that is a member of one set and not of the other, and this fact can only be due to the structure of these two sets, i.e. set membership, because set membership is the only relation of pure sets.

To be a little more formal, in these theories, structural discernibility follows from a theorem, the so-called Isomorphism Theorem.[12] As French and Krause (2006, p. 266) write, the relevant point is that $V$, the well-founded universe of ZF, is a rigid structure: in the structure described by the axioms of ZF, there are no nontrivial automorphisms; all sets are structurally discernible. French and Krause (2006, p. 265) characterize the situation in standard pure set theory as follows:

> [W]ithin the usual set-theories, indistinguishability can be considered only in relation to a certain structure, but there is

[12] See Jech (1997, p. 74). Note that the proof of the theorem proceeds by $\in$-induction, which presupposes the axiom of regularity, a.k.a. foundation.

> no indistinguishability *tout court*. That is, the objects treated by standard set theories (like Zermelo-Fraenkel with regularity) are *individuals*, in accordance with Cantor's intuitive conception of a set.

In sum, Cantor's conception of elements as discernible carries over to standard axiomatic set theories, where the elements are even structurally discernible.

### 1.4.4 Pure Set Theory and Mathematical Practice

What is the importance of pure set theory for mathematical practice? Maybe pure set theory is just a formal theory with no particular consequences for the working mathematician. However, some mathematicians think that pure set theories are the gold standard, even when it comes to mathematical practice. Halmos (1960) is an example of this line of thought. His primary concern is to provide the working mathematician with the set theoretic tools necessary for mathematical practice. Halmos gives the following account of set theory:

> What may be surprising is not so much that sets may occur as elements, but that for mathematical purposes no other elements need ever be considered. In this book, in particular, we shall study sets, and sets of sets, and similar towers of sometimes frightening height and complexity – and nothing else. By way of examples we might occasionally speak of sets of cabbages, and kings, and the like, but such usage is always to be construed as an illuminating parable only, and not as a part of the theory that is being developed. (Halmos, 1960, pp. 1)

Thus, even if only those portions of set theory necessary for mathematical practice are studied, pure set theory is entirely sufficient, if we are to believe Halmos.

### 1.4.5 Set Theory with Urelements

The crucial difference between pure set theories and set theories with urelements for our concerns is that as soon as we accept urelements, sets need no longer be structurally discernible as sets, instead the discernibility relations between urelements come into play.[13] To give an example, if we admit finitely many natural numbers as urelements, these natural numbers would not be discernible using the axiom of extensionality, but would be using the natural number structure.

Despite the fact that some mathematicians think that we need not take sets with urelements seriously, there could be good reasons for doing set

[13] See French and Krause (2006, sec. 6.4), for a more detailed discussion.

theory with urelements. Urelements are not excluded by Cantor's characterization, and if they have an important role in axiomatic set theory, then we should accept the verdict for the use of set theory in practice. I will now review mathematical and philosophical reasons for including urelements in set theory.

First, to the mathematical reasons for considering set theories with urelements. In Jech (1997, pp. 197), we find a discussion of ZF with urelements - or atoms - (ZFA) in the context of arguments for the independence of the axiom of choice (AC) from the other axioms. Jech writes that these arguments were of importance in the pre-forcing era, that is, before the independence of AC of ZF was proven using the technique of forcing. ZFA gives rise to examples where AC is violated, but, as Jech notes on page 201, these do not give any information about "true" sets, like real numbers, sets of real numbers, etc., since those sets are in $V$, the universe of ZF. Jech thinks that the importance of urelements is limited to historical and pedagogical contexts. There are no compelling mathematical reasons for including urelements in the study of (axiomatic) set theories.

Second, to the philosophical reasons for using set theories with urelements. Michael Potter (2004) proposes a version of set theory with urelements, and addresses the advantages of this approach: admitting urelements into set theory, while nonstandard, is "central for ensuring its applicability" (Ibid. p. 76). If we were to work exclusively with pure sets, "it begins to seem miraculous that mathematics applies to the world at all" (Ibid. p. 77). Potter thinks that the applicability of set theory to the world should be built into set theory, and that this can at least partially be assured by using set theory with urelements.

The issue of the applicability of mathematics is an important and pressing issue, and it is certainly valuable, if an account of set theory has the means to account for applicability. However, by admitting urelements, applicability can only be ascertained in a few cases like, for example, that of relating the number of objects to the cardinality of the corresponding set. In more complex cases, it is doubtful whether the use of urelements are of any help in accounting for the applicability of mathematics – think of population ecology, see chapter 3. In these cases, application proceeds by relating the relevant mathematical structures to the world – without detour to the (set-theoretic) domain on which the structure is defined. Unfortunately, Potter is silent on this issue.[14]

Potter is aware that set theory and application could be kept apart even in the simple cases that can be accounted for by using set theory with urelements, but he questions whether this would make sense, because at some point a correspondence between pure sets and the world has to be established:

[14] See also Moschovakis (2006, 12.35) on this issue.

> If for some reason we were determined to study only pure theory, all would not be lost quite yet, it is true: we could try to repair the damage later by adding appropriate bridging principles connecting the pure sets of our theory to the denizens of the real world that we want eventually be able to count. But it is very hard to see what the point would be of proceeding in this fashion. (Ibid., p. 77)

I think there are good reasons to separate set theory and its application. If any kind of urelement is admitted into set theory, then set theory inherits the properties of these urelements. For example, the question of the discernibility does not have a principled answer if urelements are admitted, but it has a principled answer, as we have just seen, if we consider pure set theories. It could be a good methodological decision to exclude urelements because one does not want to take their nature into account when doing set theory.

Also, if there is no restriction on the kind of admissible urelements, then set theory is not formulated in the spirit of its original conception. As we have seen, Cantor thought of the elements of sets not as just distinct, or nonidentical, but as "well-distinguished", which can reasonably be interpreted as discernible. However, if absolutely indiscernible atoms were added as urelements to ZF, this requirement would be violated.

### 1.4.6 Summary: Discernibility in Set Theory

We have now considered some important conceptions of sets, and answered the question whether, on these conceptions, elements of sets are discernible. The verdict is that on the original conception of sets, the elements are discernible, and in pure (axiomatic) set theory, elements are structurally discernible as sets. In set theories that admit urelements, discernibility depends on the discernibility of the urelements. However, we also saw that at least some mathematicians do not believe that we need urelements in set theory. Overall, there is a tendency to conceive of the elements of sets as discernible.

Of course, there is no guarantee that this tendency carries over to sets as used in mathematical practice; clearly, mathematicians do not work with pure set theory, and therefore, in practice, the elements of the sets they use are not, strictly speaking, structurally discernible. On the other hand, the case can be made that mathematicians tend to conceive of elements as discernible, and this confirms the idea that we use sets as coordinate systems, as pointed out above.

Assuming that the elements of a set are discernible, structurally or otherwise, cannot be the end of the story, because at this point, a form of Benacerraf's dilemma strikes back: if we use sets as in one of the above conceptions, say, pure sets, and we use these sets to define other structures, such as the natural numbers, the sets representing the natural numbers have "too many properties"; the sets representing natural numbers have set-theoretical properties, while natural numbers do not.

On the other hand, I argued before that we may need *some* additional properties in the domains of structures, because the domains serve as coordinate systems, and the structure that is naturally available in, say, the cardinal-four structure cannot play the role of coordinate system on its own.

The least we can expect is that, if we are given a set-theoretic representation of a structure, we get a systematic account of how to distinguish between the properties that are "really" part of the structure represented, and the properties that are only due to the representational role of set theory – we want to know what properties can be found in the landscape, and which properties are only in the map. Such an account, I think, is provided by the concept of isomorphism types.

## 1.5 Structures as Isomorphism Types

We can use the concept of isomorphism type to flesh out the idea that, on a set-theoretic construction, some properties of the structure are properly structural, while other properties are artifacts of the particular domain chosen for the representation. We have already seen an example of this idea above: cycle types capture the structural properties of cycles, while abstracting away from the particular set of labels in the domain that is permuted. Here we will consider a second example, the isomorphism type of graphs. As I pointed out above, Shapiro (2008, p. 287) argues that certain graphs are identical to, or isomorphic to, his finite cardinal structures; specifically, graphs with $n$ vertices and no edges. A further analysis of how we can reason about graphs will shed further light on the differences between a set-theoretic approach and *ante rem* structuralism.[15]

### 1.5.1 Graphs

Graphs are structures studied in pure mathematics, with applications in pure and applied mathematics, computer science, physics and other disciplines. A graph is formally defined as a pair $G = (V, E)$, with $V$ the set of vertices, and $E$ the set of edges, given as two-element subsets of $V$.[16] An example of a graph is $G = (\{a, b, c, d\}, \{\{a, b\}, \{c, d\}\})$; see a diagrammatic representation below in figure 1.4. In the following, we use $xy$ as shorthand for the two-element subset $\{x, y\}$.

[15] I thank Hannes Leitgeb for helpful correspondence on the issues discussed in this section.

[16] See e.g. Diestel (2006), a standard textbook on graph theory. I will restrict attention to finite graphs without multiple edges, without loops, and without directed edges.

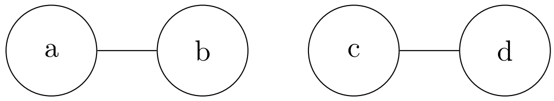



Figure 1.4: Graph $G$

### 1.5.2 Isomorphism Types of Graphs

In many cases, we do not care about the particular set $V$ used in the definition of the graph: we may not want to distinguish between the two graphs $G = (\{a,b,c,d\},\{ab,cd\})$ and $G' = (\{1,2,3,4\},\{12,34\})$ – they have the same structure, and only differ in the choice of the domain $V$. $G'$ can be obtained from $G$, by systematically substituting 1 for $a$, 2 for $b$ and so on. This is what an isomorphism between $G$ and $G'$ does: exchanging labels while leaving the structure untouched. Formally, two graphs $G = (V,E)$ and $G' = (V',E')$ are isomorphic, if there exists a bijection $\phi : V \to V'$ such that $xy \in E \iff \phi(x)\phi(y) \in E'$. The idea that we may not want to distinguish isomorphic graphs leads us to consider the *isomorphism type*, the "structure of graphs", which is the set, or class, of all graphs $G'$ isomorphic to $G$, written $[G] = \{G' : G' \simeq G\}$.

### 1.5.3 Graph Properties

Based on these notions, we can draw a distinction between properties that are structural, and properties that a graph has in virtue of the copy, or instance, we are working with. A *graph property* is a property that only depends on the isomorphism type of a graph.[17] To give an example, it is a graph property of $G$ in figure 1.4 that it is not connected, i.e. that there is no path from one component to the other. On the other hand, it is not a graph property that there is an edge between $a$ and $b$: if we use a different domain, then there may be no edge between $a$ and $b$, simply because these two elements are not in the domain of the graph.

### 1.5.4 Naming Vertices?

Using the concept of structural property, we can give a nuanced answer to the question as to whether we can name places of graphs. The answer depends on the perspective we adopt. First, we can adopt the *in re* perspective of a representant $G$ of $[G]$. As the domain $V$ is a set, its elements will usually be discernible in some sense, and thus it is usually no problem to name

[17] More generally, a function assigning equal quantities to isomorphic graphs is called a *graph invariant*. Properties are a special case of graph invariants: we can interpret them as functions from graphs to $\{1,0\}$; either a graph, and thus all graphs isomorphic to it, have a property (1), or not (0).

them. The second, *ante rem* perspective is based on the notion of graph properties of $[G]$. Let $a$ be a member of $V$. It is not a graph property of $[G]$ that $a$ is one of its vertices: there are structures $G'$ isomorphic to $G$ such that $a$ is not in the domain of $S'$. Thus, if two places in a graph are structurally indiscernible, and we only want to draw on graph properties to refer to vertices, it can happen that naming vertices is impossible.

Note the difference between this proposal and the *ante rem* structuralist's view. If we work with isomorphism types, it is not impossible to name places in a structure *tout court*. It is only impossible when we adopt a certain perspective on a structure, that is, if we restrict attention to graph properties. Other properties, namely those inherited from the domain of the structure, are still there, but they are ignored. The *ante rem* structuralist, on the other hand, maintains that the indiscernibility is an ontological feature of the structure, because he adopts an ontological reading of graph properties: these are not only the properties that are essential to the graph, but they are the only properties there are. This, I argued above, is problematic.

### 1.5.5 Functions on Isomorphism Types

It is, I think, advantageous to adopt a dual *ante rem* and *in re* perspective, if we return to the problem of how to define a function on an isomorphism type, as this construction essentially depends on the *in re* perspective.

If we want to define a function on an isomorphism type $[G]$, we first have to define a function on a *representative* of $[G]$. A representative is some member of $[G]$, i.e. one of the isomorphic copies of $G$. In a second step, we show that the function is *well-defined*. This means that the function does not depend on our choice of representative in the first step. If a function is well-defined in this sense, it is a function on the corresponding isomorphism type.

Examples of not well-defined functions can come up when one uses natural numbers as a domain, and some arithmetical expression to define the function. The problem is that these functions rely on properties of this particular domain, i.e. the fact that the elements of the domain are natural numbers. On the other hand, if I define a function on natural numbers case-by-case, the specification does not rely on the fact that the elements are natural numbers, and it will be well-defined.

As with graph properties, the trick is to first define a function on an instance of the structure, and then make sure that the function does not depend on "inessential properties" of the copy. Note that it is not clear how we would go about defining a function directly on the isomorphism type itself instead of taking a detour via an isomorphic copy. We need the "inessential properties" to get an expression for the function in the first place; only then can we make sure that the function does not really depend on these "inessential properties".

### 1.5.6 Isomorphism Types: Problems

The concept of isomorphism type is not universally accepted as an official counterpart of structures.[18] There are at least two kinds of problems.

The first worry is, roughly, that if we define isomorphism types using some fixed set theory, such as ZF, i.e. if the domain can be any set in ZF of the right size, then some of the isomorphism types will be proper classes, and thus will transcend the scope of set theory.

I think this worry is entirely justified as a foundational question about how to treat set theory itself as a structure. I do not pretend to have a solution to this kind of problem.[19] However, we know how to solve this problem for all practical purposes – i.e. all cases that do not involve set theory itself. All we need to do is to restrict the scope of elements we use for the domain of structures to a subset of the set-theoretic universe. For example, an initial segment of the natural numbers is entirely sufficient as a "background ontology" for all finite structures. This blocks the problem that isomorphism classes are proper classes.

The second worry we can have about structures as isomorphism types goes to the core of the *ante rem* structuralist's problem with set-theoretic structuralism. It is formulated in Leitgeb and Ladyman (2008) for the case of graph theory. Leitgeb and Ladyman think that isomorphism types of graphs do not really capture what graphs theorists are after when they reason about graphs, because isomorphism types do have features that are absent in structures, and vice versa: graphs don't have members, and sets don't have edges.

I think that this is a valid point. We should not *identify* graphs, or any abstract structure, with isomorphism types. Isomorphism types have, as I explained, two aspects: a representational, or *in re* aspect that is rooted in the domain of structures, and a non-representational, or *ante rem* aspect that only takes the "essential structure", or graph structure, into account. Leitgeb and Ladyman are right in pointing out that isomorphism types have purely representational "surplus structure".

However, the *graph properties* are exactly what the graph theorist is after – they are the properties that are essential to the graph. I think Leitgeb and Ladyman would agree. The crucial point of disagreement is that on their view, different vertices of graphs can be indistinguishable *simpliciter*. I argued that this is problematic: We need the "surplus structure" in order to define functions on isomorphism types; graph properties are simply not sufficient do that job. Both the *ante rem* and the *in re* perspective that isomorphism types provide are an indispensable part of many mathematical structures.

[18] See e.g. Hellman (2005, p. 539) and Shapiro (1997, p. 92).
[19] See Hellman (2005) for an overview of corresponding problems.

### 1.5.7 Conclusion

Summing up, isomorphism types seem to capture a good part of what we want from mathematical structures. On the one hand, there is no problem with establishing correspondences between isomorphism types or defining functions on isomorphism types, as we can adopt an *in re* perspective; on the other hand, we can also consider the "purely structural" aspects of an isomorphism type, for example the graph properties, that have nothing to do with the set-theoretical construction of graphs.

However, we also have to acknowledge that isomorphism types do not capture all aspects of mathematical structures. First, there is the foundational problem of accounting for set theory itself in structural terms. Second, there is no denying that isomorphism types have properties that we would not want to attribute to, say, graphs.

The second point appears to raise a dilemma that does not seem to be due to set theory in particular, but to mathematical practice in general. We would like to characterize our structures in terms of their proper structural properties and relations. This blocks our ability to discern places in some structures, which is necessary to define functions on these structures. This is the *ante rem* horn. However, if we accept that our structures are characterized in set-theoretic terms, our structures inherit surplus structure, and the set-theoretical representations have properties that we would not normally attribute to our structures. This is the *in re* horn.

While I do not see a completely satisfactory solution to this dilemma, I think we should take the *in re* horn, for the following reason. If we choose this option, it is at least possible to adopt both an *ante rem* and an *in re* perspective on a structure: we can consider the "essentially structural", *ante rem* properties of a structure. What we do not get in this way is an *ante rem* ontology.

A passage in Shapiro (1997, p. 74) suggests that the basic idea behind *ante rem* structuralism might be compatible with isomorphism types after all:

> A *structure* is the abstract form of a system, highlighting the interrelationships among the objects, and ignoring any features of them that do not affect how they relate to other objects in the system.

It seems as if in this passage, Shapiro does not insist on a reification of the *ante rem* perspective of structures: if the "ignored features" (representational properties in the domain) are not completely eliminated from the structure, but are just distinguished from the "highlighted interrelationships" (structural properties), then *ante rem* structures are compatible with structures as isomorphism types.

# Chapter 2

# The Bridges of Königsberg

## 2.1 Introduction

In this chapter, I reexamine Leonhard Euler's solution of the Königsberg bridges problem, a case of the application of mathematics that has become standard in the pertinent philosophical debates. The problem Euler faced was to determine whether or not it is possible to cross every bridge in the ancient city of Königsberg exactly once. Euler solved the problem – there is no such path – by relating it to a mathematical problem, to which he proposed multiple solutions. Since its introduction into the debate by Torsten Wilholt (2004), philosophers have been interested in this case because Euler's solution can be interpreted as an explanation of a scientific phenomenon in which mathematics plays an important role, and it has been used to propose novel kinds of scientific explanations.

This chapter improves on previous philosophical accounts by basing the discussion on Euler's original paper on the case, as well as other historical sources; see section 2.2. This has several advantages. Euler proposes not one, but at least three different solutions to the problem. His discussion of the respective strengths and weaknesses of these solutions can be fruitfully explicated in terms of differences in explanatory power; see section 2.3.5. Also, we can retrace the genesis of Euler's solution, on the basis of historical sources. This allows us to better understand what is involved in the "mathematization" of the problem, i.e. the transformation of a problem through the use of mathematics.

In section 2.3, after introducing some conceptual distinctions, I propose interpreting Euler's solution both as an explanation within mathematics and as a scientific explanation. The purely mathematical explanation is not a proof, but an application of Euler's theorem. Thus, explanations within mathematics need not be proofs. I put this notion of intra-mathematical explanation to work against two recent philosophical accounts of the Königsberg case, by Alan Baker (2012) and Marc Lange (2013), in section 2.4.

In section 2.5, I first discuss whether the applied version of Euler's solution can be interpreted as a causal explanation. I suggest that while this is not a causal explanation in the narrow sense that it explains by giving a cause, we can give the components of the underlying mathematical structure a causal reading. Thus, on a broad reading of the notion, this is a causal explanation. Finally, I claim that a pragmatist reading of the relation between the mathematical formulation of the problem and its real-world counterpart, based on Bas van Fraassen's theory of explanations, allows an adequate understanding of the case.

In section 2.6, I examine two philosophical analyses of the Königsberg case, in view of the preceding analysis. I argue that the proposal by Christopher Pincock (2007) to reconstruct the case as an "abstract explanation" is incomplete, and that the account of "distinctively mathematical explanations" by Marc Lange (2013), which goes against a causal reading of the case, conflates the applied and the purely mathematical version of Euler's solution.

The following picture of the role of mathematics emerges. We can discern two different kinds of contributions of mathematics to Euler's explanations. First, mathematics contributes to explanatory power on the level of pure mathematics. Characteristics of explanatory power are reduction of irrelevant information, reduction of complexity, and increased simplicity. Second, the mathematics aids explanations by representing aspects of the causal structure of the city of Königsberg. This is a causal explanation on a liberal notion of causal explanation, if we take the pragmatics of explanations into account.

## 2.2 Euler's Königsberg

In this section, after a short look at the genesis of the Königsberg bridges problem, I scrutinize Euler's original solution to the Königsberg bridges problem in his paper "Solutio problematis ad geometriam situs pertinentis" ("The solution of a problem relating to the geometry of position"), and I compare Euler's approach to a modern approach.[1]

### 2.2.1 Prehistory

The story of how Euler learned of the Königsberg bridges problem is not completely known, see Sachs et al. (1988). Euler probably first heard about the problem from letters by Carl Leonhard Gottlieb Ehler, who acted as an intermediary between Euler and Heinrich Kühn, a professor of mathematics

[1] I use the widely available translation Euler (1956). See Hopkins and Wilson (2004) for a useful overview of Euler's paper. I thank an anonymous referee for his suggestion to consider Euler's original publication, and for pointing out several interesting aspects of Euler's approach.

from Danzig. In a letter to Euler on March 9, 1736, Ehler alludes to an earlier formulation of the problem, asks for a solution, and adds a schematic map of the city of Königsberg, indicating the direction of flow of the river and stating the names of bridges, the island, and neighborhoods.

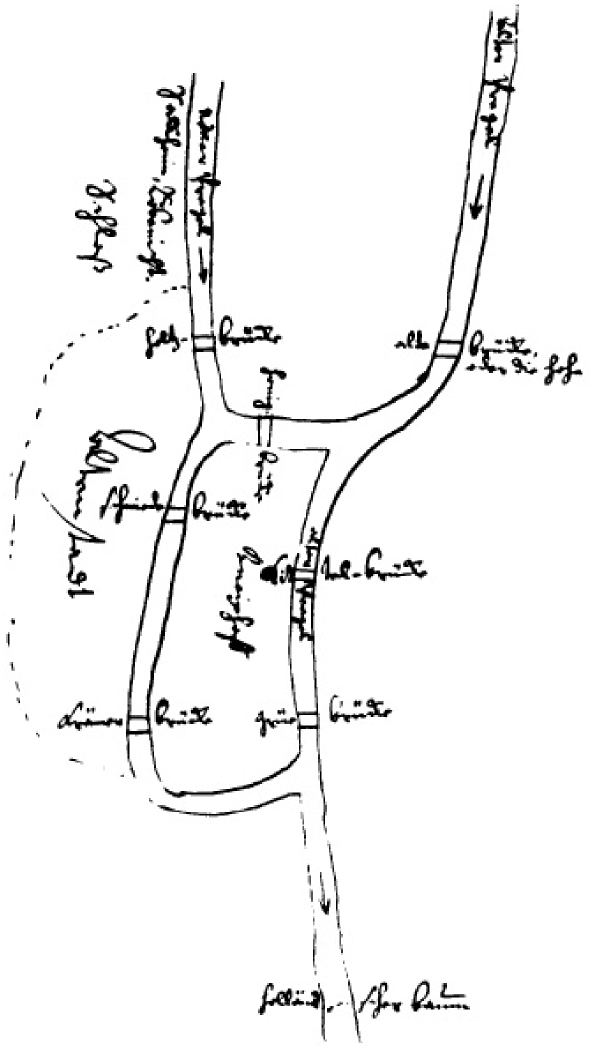


Figure 2.1: Ehler's Map of Königsberg

The solution to the problem is laid out in full detail in the Königsberg paper. Euler probably obtained the solution a few days after receiving Ehler's letter; he noted that he had a solution in a letter to Giovanni Jacobo Marinoni on March 13, 1736. However, Euler had not yet made up his mind as to the mathematical significance of the problem. In a letter to Ehler on April 3, 1736, he wrote that "the solution is based on reason alone, and its discovery does not depend on any mathematical principle"; he also had not yet assigned the problem to a new mathematical discipline: "[Y]ou have assigned this question to the geometry of position, but I am ignorant as to what this new discipline involves [...]". Reading the Königsberg paper shows that Euler would soon make up his mind about these matters.

### 2.2.2 The Königsberg Paper

We will now reconstruct Euler's line of thought in the Königsberg paper.

In paragraph 1, Euler states his systematic interest in the Königsberg bridges problem. He takes it to be an example of a new, special kind of geometry, which does not involve quantities and measures, but only position. Leibniz introduced this new kind of geometry under the name *geometria situs*, geometry of place, what we would now call topology. Euler writes that no systematic account of this kind geometry is available yet, but that the problem at hand can serve to illustrate the new theory and its methods. Euler apparently changed his mind about the mathematical nature of the problem and its classification as topological.

In paragraph 2, Euler describes two problems. He illustrates the situation in Königsberg using a schematic map; see figure 2.2. He assigns capital letters $A, B, C, D$ to the island and land masses, and lower case letters $a, b, c, ...$ to the bridges connecting land masses and island. The first problem is to find out whether it is possible to cross every bridge of this system exactly once. Euler notes that there is no definite answer to this problem as yet. We will call this the *Königsberg Problem*. Euler then generalizes the problem and asks how one can determine the solution, not only for this particular configuration, but for any kind of system, i.e. any kind of branching of the river and any number of bridges. We will call this second problem the *General Problem*.

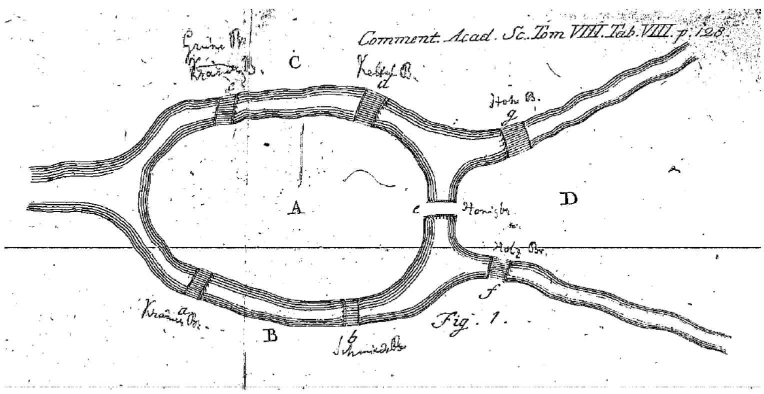



Figure 2.2: Euler's Map of Königsberg

Several aspects of this paragraph are noteworthy. Firstly, there is the introduction of two kinds of letters, one for places, or areas, the other for connections, or bridges. This distinction, as we will see in a moment, is key to a graph-theoretic approach to both problems. This notational innovation does not feature in Ehler's letter of March 9, 1736. Secondly, Euler suppressed some information contained in Ehler's map, such as the names of bridges and the direction of flow of the river, probably because he considered them to be irrelevant to the solution. Third, the immediate generalization of the problem shifts the focus from mere puzzle-solving to a real mathematical question – some solutions that are appropriate for the special case of Königsberg will not do in the general case, which requires a more elegant argument. The Königsberg Problem and the General Problem should be carefully distinguished.

In paragraph 3, we learn of a first method for solving the Königsberg Problem: it consists of "tabulating all possible paths" and examining whether one of them uses every bridge exactly once. Euler rejects this method because it is "too tedious and too difficult": there are too many possible paths, and for bigger systems, this method becomes intractable. The approach generates information that is irrelevant to the problem at hand. This insight leads to a shift of focus: Euler restricts the task to establishing whether the required path exists, which does not require the specification of an actual path, i.e. a witness. Euler does not specify how to carry out the brute force search in detail. However, he thinks that it will be a finite procedure.[2]

A crucial innovation is introduced in paragraph 4. It consists of the use of a particular notation for paths in the bridge system in terms of the crossing of bridges. All bridges are labeled with lower-case letters $a, b, c, ...$; see figure 2.2. Land masses and islands, on the other hand, are labeled with capital letters $A, B, C, ...$ . The crossing of any one of the bridges $a$ and $b$ between $A$ and $B$ can now be written as $AB$. A path from $A$ over $B$ to $D$ is noted as $ABD$, using any one of the bridges connecting these areas.

This notational shift is characteristic of the graph-theoretic nature of Euler's approach to both problems and marks the invention of graph theory. In modern graph theory, a graph[3] is represented by a set of vertices $V$, and a set of edges $E$, represented by pairs of vertices, $E \subseteq V^2$. This is exactly what Euler's notation achieves: Bridges and areas are brought into notational correspondence by writing bridges (edges) as pairs of areas (vertices). There are no longer two separate sets of labels for the two kinds of objects – bridges (edges) are expressed with the help of areas (vertices).

As Euler himself notes, the bulk of the paper consists of putting this

[2] One way of implementing it would be to write down all (finitely many) paths of length seven starting from any one of the areas $A, B, C, D$, and see whether any one of these paths consists of seven different bridges. A distinctive feature of the brute force approach is that we do not need both kinds of letters introduced above – the lower-case bridge labels will do the job.

[3] More specifically, a multigraph, as there are some pairs of vertices that are connected by more than one edge.

simple yet powerful idea to work. The rest of the paper is a sequence of methods for determining the existence of what we now know as Euler paths, culminating in Euler's theorem. The key idea is that if we write bridges as pairs of areas, we can compare the algebraic condition for the length of Euler paths with algebraic conditions for the areas.

First, Euler finds a method that is sufficient to solve the Königsberg Problem. We can represent a path on a bridge system consisting of $n$ bridges by a string of $n+1$ capital letters, i.e. land masses. A bridge between land masses $A$ and $B$ is written $AB$. The notation does not distinguish between different bridges that connect the same two land masses, as this information is irrelevant for the existence of a path. However, we know that if we want to use every bridge in Königsberg exactly once, the string has to consist of 8 letters.

We can now figure out how many times a capital letter (area) has to occur in a string representing a path. If the number of bridges leading to area $X$ is odd $(2n-1)$, then $X$ will have to appear $n$ times: If three bridges lead to area $X$, then $X$ will have to feature twice in the path, whether we start in area $X$ or not. For five bridges, the letter has to occur three times, and so on. We can now apply this result to the Königsberg system. We need $A$ three times, and $B, C, D$ two times. This adds up to 9, which is bigger than 8. It is therefore impossible to find a path that crosses every bridge in Königsberg exactly once.

Euler's next step is to extend this method to systems with even areas. If an area $X$ has an even number of bridges $(2n)$, there are two possibilities. If we start in $X$, the letter $X$ will occur $n+1$ times. If we do not start in $X$, it will occur $n$ times. We can now sum up the number of times a letter has to occur in a path string for even and odd regions – taking the starting point into account – and compare the result with the length of an Euler path. This is a solution to the General Problem. If the sum is equal to the number of bridges plus one, then there is a path, but only if we start in an odd region. If the result is less than that, we can start in an even region.

After explaining this method in some detail, and illustrating with a more complicated, imaginary bridge system, Euler writes:

> By this method we can easily determine, even in cases of considerable complexity, whether a single crossing of each of the bridges in sequence is actually possible. But I should now like to give another and much simpler method, which follows quite easily from the preceding [...] (Euler, 1956, par. 16).

This "much simpler method" yields Euler's theorem. It is based on the preceding method, together with the observation that if we separately determine the number of bridges adjacent to each region, and add these numbers up, we get double the actual number of bridges, because each bridge is counted twice. The sum thus has to be even. The theorem distinguishes the following cases: If all regions are even, it is possible to cross all the bridges no matter where we start. If two regions are odd, there is a path, provided that we start in an odd region. However, if four, six, etc. regions are odd, their sum is greater than the number of bridges plus one, thus no Euler path exists.

### 2.2.3 Euler and Modern Graph Theory

Here I compare Euler's solution to modern formulations of the Königsberg Problem and the General Problem.[4]

The most important difference is that many modern formulations heavily rely on graph diagrams, such as in the following figure, to explain the problem.

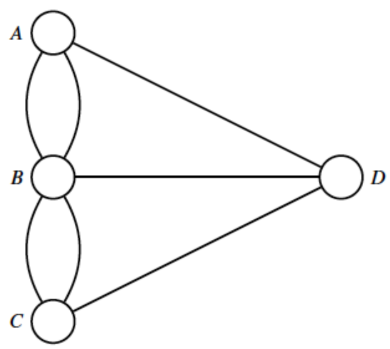



Figure 2.3: Königsberg Graph

However, no such diagram can be found in Euler's paper. He only uses various schematic maps of Königsberg, see figure 2.2, and of other, imaginary bridge systems. As Robin J. Wilson (1986, p. 272) points out, graph diagrams only appear 150 years later.

The fact that Euler's reasoning does not rely on a graph diagram does not mean that it is not graph theoretical. The omission of a diagrammatic representation probably has historical reasons. According to Kruja et al. (2002), the omission can be traced back to the influence of Leibniz, who envisioned the new "geometria situs" as dispensing with figures which, to his mind, hindered the imagination. I argued above that the introduction of graph-theoretic methods in the paper occurs in paragraph four with the identification of bridges (lower case letters) with pairs of land masses (pairs of capital letters). This is the relevant innovation, not the use of graph diagrams.

[4]For a modern textbook account, including a proof of Euler's theorem, see Diestel (2006, pp. 21).

Further differences between Euler's and the modern approach should be noted. First, Euler's initial question is slightly different from the modern formulation: he asks if it is possible to cross every bridge in Königsberg exactly once, without the further assumption that starting and end point coincide. Second, Euler does not assume that graphs are connected. This is a necessary condition for the theorem: if a graph is not connected, we will not be able to find an Euler path even if all vertices have even valence, simply because at least one part of the graph is inaccessible from another. Euler probably omitted this condition because it is implicitly clear that there is no solution in these cases. Third, Euler only proved one direction of the eponymous theorem. The theorem states that an Euler path on a connected graph exists if, and only if, every vertex has an even number of edges. The fact that if an Euler path exists, every vertex has an even number of edges was only proved 135 years later; see Wilson (1986, p. 270). Note that this direction of the theorem is not necessary for the solution of the Königsberg Problem.

## 2.3 Königsberg Within Mathematics

We now turn to a philosophical analysis of Euler's work. The focus of this section is on explanations within pure mathematics; the discussion of the Königsberg case as a scientific explanation follows in section 2.5. It will turn out that it is fruitful to interpret the differences between Euler's methods to solve the Königsberg and the General Problem in terms of differences in explanatory power.

### 2.3.1 Intra-Mathematical Explanations and Scientific Explanations Using Mathematics

The discussion on mathematical explanations distinguishes between mathematical explanations of purely mathematical phenomena, and scientific explanations that make use of mathematics.[5] Both kinds of explanation are controversial and have been hotly debated in recent years: some examples of Intra-Mathematical Explanation (IME) have been put forward both by philosophers and mathematicians, but we do not yet have a stable notion of what IME involves. We certainly need more good real-life examples. Scientific Explanation using Mathematics (SEM), on the other hand, is under debate, mainly because it has been used to defend mathematical Platonism in the context of indispensability arguments.[6]

[5] The distinction is due to Baker (2012). See Mancosu (2011) for a useful overview of the debate on explanations in pure and applied mathematics.

[6] See Baker (2009) for the now-standard formulation of an explanatory version of indispensability arguments. I will not discuss whether Euler's solution to the Königsberg bridges problem speaks in favor of mathematical Platonism.

*Prima facie*, the distinction between IME and SEM is not particularly controversial. All we need is a clear separation between an empirical and a mathematical domain, and at least some idea of what an explanation involves. The distinction between the two domains is not hard to draw in the present case: some parts of Euler's argument only concern pure graph theory and are unrelated to any particular domain of application, while other parts allow us to draw conclusions about the existence or non-existence of paths in cities, and Königsberg in particular. It will be instructive to pay close attention to this difference. We will see in the discussion of recent philosophical work that these notions are still debated.

As to the explanatory nature of Euler's work, I will only presuppose that it is at least partially geared towards answering why questions. My focus is on the question as to what we can learn from Euler's work if we approach it from an explanatory perspective, and what the distinctive role of mathematics in explanations is.[7]

### 2.3.2 The Königsberg Problem and the General Problem

In paragraph 2, Euler distinguishes between the question whether there is a certain path in the city of Königsberg, the *Königsberg Problem*, and the general question of conditions for paths in similar kinds of system, the *General Problem*. Both problems give rise to explanations. In the case of the Königsberg Problem, the relevant why question is: why is it impossible to cross every bridge in Königsberg exactly once? In the case of the General Problem, the question is: why is it possible to cross every bridge in some systems (of a certain kind), and not in others? The two problems are not independent: the General Problem also solves the Königsberg Problem as a special case.

Do the explanations that answer these two questions belong to pure mathematics (IME), or are they scientific explanations that come with a portion of mathematics (SEM)? It seems to me that we can construe both explanations as belonging to both kinds of explanations: they can concern one real (or several actual and possible) bridge systems, or they can concern one (or a type of) graph-theoretic problem. I think that one of the keys to a better understanding of both IME and SEM is to get a clear picture of how the purely mathematical formulation, and the scientific formulation of the explanations, are related.

[7] I will not enter into debates about particular models of scientific explanation and how they mesh with IME and SEM; the following remarks have to suffice. First, it has been convincingly argued in the literature that mathematical explanations should not be classified as explanations by unification; see Hafner and Mancosu (2008). Second, at least IMEs cannot be given a causal reading. Therefore, explanations in mathematics do not fit any of the common models of scientific explanation.

### 2.3.3 The Brute Force Solution vs. Euler's Solution

I argued in section 2.2.2 above that the introduction of two sets of labels, and the fact that we can express one set (bridge labels) in terms of the other (area labels) marks the invention of graph theory. Euler notes early on, in paragraph 3, that we can tabulate all possible paths and examine whether one of these conforms to our specifications, and that this solves the Königsberg Problem. This, in turn, means that we do not need both sets of labels to solve the Königsberg Problem: we can tabulate paths in terms of bridge labels; there is no need for the area letters. Strictly speaking, we do not need graph theory to solve the Königsberg Problem.

However, Euler quickly dismisses this approach (Brute Force Method henceforth) as not satisfactory. His reason to dismiss the approach is that it is "tedious": It solves the problem, but not in an efficient and telling manner. It gives us too many irrelevant details and is therefore not informative and tractable. Euler is after a more telling method for solving the problem. This superior method, which I will call Euler's Method, underlies Euler's theorem, which states that an Euler path exists if and only if all vertices have an even number of edges, or if exactly two vertices have an odd number of edges.

I think it is fruitful to interpret the difference between these two methods in terms of explanatory power. Both methods provide answers to the why question of the Königsberg Problem. However, the answer provided by the Brute Force Method does not meet Euler's expectations for a solution to the Königsberg Problem. The method does give us a reason why an Euler path on the Königsberg system does not exists – none of the paths in a complete list is an Euler path – but the reason is not very telling. Euler writes, "it yields a great many details that are irrelevant to the problem".[8] The Euler method, on the other hand, provides an *explanatory* answer. It tells us that the reason for the non-existence of an Euler path is that none of the four areas has an even valence.

Both the Brute Force Method and Euler's method can be interpreted as intra-mathematical explanations (IMEs): In the case of the Brute Force Method, the (bad) explanation consists of a list of strings of lower case letters of length seven, the fact that this list contains all strings that comply with the structure of the Königsberg map, and the fact that none of these strings consists of seven different letters. In the case of Euler's Method, the explanation consists of the Königsberg graph, Euler's theorem, and the fact that according to the theorem, there is not Euler path on the Königsberg graph.

[8]We could also interpret the Brute Force Method as not being an explanation at all. However, I think this would be going too far. After all, the method does give us a reason, albeit a not very telling one. I therefore prefer to interpret it as a *bad* explanation.

### 2.3.4 Two Notions of Intra-Mathematical Explanations

Both the Brute Force Method and the Euler Method can be used to solve the Königsberg Problem. They have a similar form: both contain general facts about the method to solve the problem, and also facts about the Königsberg system. On the Brute Force Method, we use the general fact that we can compile a complete list of paths of a certain length for a system, and the particular list of the Königsberg system. In Euler's explanation, we use Euler's theorem, and the Königsberg graph.

It is important to note that in both cases, we use general facts, such as theorems, but the proofs of these facts are not part of the explanation. The mere appeal to Euler's theorem, and the fact that the Königsberg system does not satisfy the theorem's conditions, is a legitimate and satisfactory explanation. The proof of the theorem is not a necessary ingredient of the explanation. The same is true, to a certain extent, for the Brute Force explanation. The explanation is satisfactory if we accept that the Brute Force Method works, and the proof that this is so need not be part of the explanation.[9]

This is not to say that we have to accept Euler's theorem or the Brute Force Method on faith. We can request proof of these results. However, this is a request for different explanations, namely the explanations of methods or theorems. We can keep apart the explanation of a theorem, which can consist in a proof, and an application of the same theorem in the explanation of a particular *mathematical* fact. We should not be too puzzled about this, as the same is common in scientific explanations: we use a regularity to explain an event, and we can explain the regularity by appealing to a different, more general regularity. These are just two different explanations.

The difference between these two kinds of explanations is the one between a solution to the General Problem and a solution to the Königsberg Problem. The former consists of a method to decide whether a class of systems has a certain property, while the latter is the application of this method to a particular system. Euler deemed only the General Problem a genuine mathematical problem. This mirrors the different status of the two explanations. While both are deductive, only the solution to the General Problem is the proof of a theorem; the solution of the Königsberg Problem is merely an application of that theorem.

Euler's two solutions two the Königsberg Problem have a few things in common. They do not involve a proof and they both use general facts and facts pertaining to the Königsberg system. But what are the (explanatory) differences between the methods?

---

[9] The idea that we can draw this distinction is taken from a recent paper by Alan Baker (2012). We will discuss the ramifications of the distinction for Baker's ideas in section 2.4.1 below.

### 2.3.5 Explanatory Differences Between Euler's IMEs

Euler notes several differences. The first is the "large number of possible combinations" of the Brute Force Method, which makes it difficult to carry out, or even inapplicable for more complicated systems – a *Complexity Difference*. Second, Euler points out that he is after a simpler method, i.e. there is a *Simplicity Difference*. The third difference is that the Brute Force Method gives us "a great many details that are irrelevant to the problem", i.e. an *Irrelevant Information Difference*. Euler thinks that the Irrelevant Information Difference is responsible for the Complexity Difference. Fourth, if an Euler path exists, the Brute Force Method will find it. This goal is relaxed for the Euler Method. It is not necessary to specify a path; a method that decides whether a path exists or not is sufficient. Call this the *Witness Difference*. Here I will examine whether these differences shed light on the difference in explanatory power between the two methods. I do not claim that these differences are unrelated or mutually exclusive.

#### The Complexity Difference

First, to the Complexity Difference.[10] It is plausible that a reduction of complexity contributes to the goodness of an explanation. A method that reduces complexity can enhance our understanding, because it can make it easier for us to grasp the reason for the existence or non-existence of a path, given that it can be hard to survey the long list of possible paths of a complex system.

However, Euler is not only after a reduction of complexity. Before discussing Euler's theorem, he describes a different, but general method to solve the Königsberg Problem. In this "Intermediate Method", we sum up the number of times the letter has to occur in a string, depending on the number of bridges connected to an area. Euler notes that while the Intermediate Method is already quite successful at reducing complexity, he prefers a "much simpler method", which we dub "Euler's Method", culminating in Euler's theorem. This suggests that while complexity can contribute to the goodness of an explanation, it is not sufficient: the Intermediate Method already achieves a (sufficient) reduction of complexity, but it does not have a sufficient degree of simplicity.

#### The Simplicity Difference

Simplicity is a well-known candidate contributor to explanatory power. What variety of simplicity could Euler have in mind? Euler's Method is easier to

[10] I think that Euler's pre-theoretic notion of complexity could be spelled out in terms of modern computational complexity, and that it would be helpful to carry out such an analysis. However, I will not explore the computational complexity of the methods under discussion here.

carry out than the Intermediate Method, as for the latter, we have to count the number of bridges for each area, determine the number of times the area letter has to occur in a path string, sum the results up, and compare the number with the length of Euler paths of the system. If, on the other hand, we use Euler's Method, the only information we need is whether zero or two areas have an odd number of bridges. This is easier to determine than the calculation of the Intermediate Method.

Euler's Method is also simpler in that it provides us with an intuitive reason as to why an Euler path does or does not exist in a system with one or more than three odd areas. In a discussion of the case, Marc Lange (2013, p. 5) sums it up as follows: "Any successful bridge-crosser would have to enter a given [area] exactly as many times as she leaves it unless that [area] is the start or the end of her trip. So among the [areas], either none (if the trip starts and ends at the same vertex) or two could touch an odd number of edges."

The Intermediate Method provides us with a reason that is *somewhat* transparent – the method of determining the numbers for odd and even areas and that we have to sum them up makes intuitive sense – but it is not as clear as Euler's theorem, where we get an immediate grasp on the relevant property. The Intermediate Method, however, is clearer than the reason provided by the Brute Force Method, where we do not get any information about how the (non-)existence of paths depends on the structure of the bridge system.

Summing up, there could be two varieties of simplicity at play. First, the methods with more explanatory power are easier to carry out – this could be related to complexity. Second, the reason provided by the more explanatory methods becomes more transparent. This, in turn, could be related to the Irrelevant Information Difference, to which we will now turn.

### The Irrelevant Information Difference

Euler writes that the Brute Force Method gives us details that are irrelevant to the problem at hand. In one sense, this is wrong: the method does not give us any irrelevant details at all, because we need a complete list of possible paths of a certain length to establish that there really is no Euler path on the system. Yet, in another sense, Euler is right. Going through all these possibilities to figure out whether there is a path or not seems like overkill. Why?

The reason could be that every time we write down one possible path, we draw on the structure of the bridge system, because this structure dictates what sequence is possible. However, it can happen that we use the same information more than once, for example if two paths share an initial segment. The structural information of the bridge system is plugged into the paths, but in a redundant manner.

The Intermediate Method does better in this respect than the Brute Force Method. We compute the length of an Euler path in a compositional manner, by exploiting a structural property of areas, namely the number of bridges connected to it. As we use every area once, redundancy is reduced and because of the compositionality, i.e. the fact that the property of the system depends on the property of the parts in a clear manner, the information is more transparent.

On the other hand, if we compare the Intermediate Method with Euler's Method, we notice that the former still exhibits some redundancy, because it requires us to determine the exact number of bridges connected to each area, compute another number from this, and sum the results up. In Euler's Method, most of this information is irrelevant. All that matters is whether the number of bridges connected to an area is even or odd. In some sense, Euler's Method is maximally informative in that it uses no irrelevant information at all. Euler's theorem is an equivalence: it is sufficient, and necessary, to know whether the numbers are even or odd to answer the why question.

There exists a philosophical account of IME proposed by Mark Steiner (1978a) that could shed light on the Irrelevant Information Difference. Steiner suggests an analysis of explanatory proofs in terms of *characterizing properties*:

> My view exploits the idea that to explain the behavior of an entity, one deduces the behavior from the essence or nature of the entity. [...] Instead of 'essence', I shall speak of 'characterizing properties', by which I mean a property unique to a given entity or structure within a family or domain of such entities or structure. My proposal is that an explanatory proof makes reference to a characterizing property of an entity or structure mentioned in the theorem, such that from the proof it is evident that the result depends on the property. (Ibid., p. 143)

This account suggests that in an IME, the explanatory work is done by a characterizing property, a property used in the *explanans* that somehow characterizes the *explanandum.*

Let's apply this idea to the three methods. In the Brute Force Method, there is no property of the structure that we use in particular, we go directly for the possible paths, and the structural information is "evenly distributed" over the paths. In the Intermediate Method, we identify a relevant structural property, the valence of areas. This property is sufficient to account for our explanandum. Thus, the valence of the areas probably qualifies as a characterizing property: It is sufficient to pick out exactly those structures on which an Euler path exists. Finally, in Euler's Method, the characterizing property of the Intermediate Method is "refined": We only need the property of even or odd valence of areas. The combination of this property of areas in Euler's theorem gives us not only a sufficient, but also a necessary condition for our explanandum.

What is the relation between characterizing properties and irrelevant information? I suggest that characterizing properties are a means for getting rid of irrelevant information. Not all of the structural information of the Königsberg graph matters for the *explanandum* we are after. A good characterizing property squeezes exactly the right amount of information out of the structure, while a not-so-good characterizing property squeezes out too much. This suggests that characterizing properties can come in degrees. The less (irrelevant) information it contains, the better it is, the ideal case being an characterizing property that is equivalent to the explanandum.

However, caveats are in order. I do not endorse Steiner's proposal as a general account of IME. He proposes a notion of IME that focuses on explanatory proofs. I suggested above that we should distinguish two notions of IME, one concerning explanations of theorems, which are typically proofs, the other explanations that consist of the application of a theorem to a particular mathematical fact or structure, which need not involve proofs. It is not clear that Steiner's proposal captures both notions equally well.

Steiner's proposal has been criticized on several occasions; see Resnik and Kushner (1987), and Hafner and Mancosu (2005). They show that Steiner's account is not all there is to IME in proposing examples that are not easily captured on Steiner's view. However, I think it would be hasty to read these objections as overall rejections of Steiner's account. Maybe it is too early in the discussion on IMEs to try to formulate a completely general account of these explanations, or maybe IMEs are inherently heterogeneous. At this stage of the debate, we should not focus too much on the old game of rejecting overreaching philosophical accounts with counterexamples, but rather try to come up with plausible candidates that can shed light on, at least some, real-life examples.

In sum, the idea that characterizing properties are relevant, in that they are a "measure" of the amount of irrelevant information, and thereby correlated with explanatory power, is not too far-fetched.

### The Witness Difference

The Brute Force Method, if applied to a system that has an Euler path, would provide us not only with one specimen, but with all Euler paths in that system, while both the Intermediate Method and Euler's Method do not provide us with this information. In a sense, this is a deficiency of those two methods that we would otherwise assess as more explanatory. There are two ways in which we can deal with this problem. We can either insist that we are only interested in the existence of an Euler path, i.e. in a decision method, and not in witnesses – the request for a witness is a different explanandum.

Alternatively, we can modify the two methods so as to produce witnesses.

In the last paragraph of the paper, Euler writes that in cases where an Euler path exists, we are left with the task of finding such a path. He then describes a method that facilitates the search: Eliminate any two bridges that connect the same region, and find a path on the reduced system. This should be considerably easier than the original task. If this is done, we can just add the omitted pairs of bridges in "loops" to the reduced systems, which results in an Euler path in the original system.

Euler thus chooses the second option. This supports the view that it is a virtue of the Brute Force Method to deliver candidate Euler paths if they exist. The Brute Force Method has at least one explanatory virtue after all.

## 2.4 Philosophers on the Transmission View

The above discussion sheds light on some issues in the recent philosophical debate on scientific explanations using mathematics. Here we will take a look at contributions by Alan Baker (2012) and Mark Lange (2013). Both are sceptical of a certain view of IME and its application. I argue that their scepticism is difficult to maintain in view of the above discussion.

### 2.4.1 Baker on the Transmission View

Baker's arguments are directed against a view of SEM defended by Mark Steiner (1978b), which Baker dubs the *Transmission View*. Here is a short reconstruction of this view (see chapter 4; it is repeated here for readability).

According to the Transmission View, SEM work via a transmission of an IME to some physical explanandum. The SEM with explanans $M$, typically a proof, used in the explanation of a physical explanandum $P^*$, written $M \rightarrow P^*$ is, first and foremost, an explanation of an IME $M^*$, written $M \rightarrow M^*$, and the explanation of $M^*$ is transmitted to $P^*$ via a bridge principle, written $M^* \leftrightarrow P^*$. If we remove the bridge principle from the complete SEM, $M \rightarrow M^* \leftrightarrow P^*$, we are left with an IME, $M \rightarrow M^*$.

Baker identifies two separate problems with this view. The first is a counterexample, the honeycomb case. I deal with this example *in extenso* in chapter 4. The second is an argument for the thesis that, even if we use a theorem in an SEM, the proof of said theorem is not necessarily part of the SEM.

Baker supports his argument with a well-known case of SEM, the cicada case.[11] The *explanandum* is that certain cicadas have live cycles with periods of 17 years (or some other prime-numbered period). The *explanans* consists of biological and mathematical premisses. The biological premisses are that

[11] Here we follow the cicada case as introduced in Baker (2005); see also the discussion and reactions to critics in Baker (2009).

it is evolutionary advantageous to minimize intersection with other periods (to avoid predators emerging in periods, or to avoid hybridization), and that there are ecological constraints on the possible length of life cycles (they have to lie between 14 and 18 years). The mathematical premiss, a number-theoretic theorem, is that prime periods minimize intersection. From this we can deduce that cicadas will have a life cycle of 17 years.

Baker argues that in this explanation, a number-theoretic theorem is used, but the proof of said theorem does not feature in the explanation. From this he concludes that the Transmission View is flawed, as the standard explanations of the cicada case do not use the proof of the number-theoretic theorem. According to him, it is sufficient for the explanation *that* the theorem has been proved.

I agree with Baker that in this example, the proof of the number-theoretic theorem need not be part of the explanation. However, this only undermines the Transmission View if all IMEs are, or involve, proofs. I argued above that the application of Euler's theorem to the Königsberg graph constitutes an IME such that no proof features in the explanation. The explanation is more like a corollary, or an application of the theorem within mathematics.

Baker makes the implicit, but important, presupposition that the explanatory relation within mathematics has to involve a proof. If this is not so, we can accept that the cicada case is a good explanation without using the proof of the number-theoretic theorem, and still defend a version of the Transmission View.

I do not want to defend the Transmission View as a general account of mathematical explanations in science. However, it seems to have certain advantages that are worth keeping in mind. For one, according to this view, mathematics can have explanatory benefits both within mathematics (due to the IME part of the explanation), and in application to the world (due to the bridge principle). It also seems to fit nicely with at least some candidates for SEM, one of them being the Königsberg case.

On the other hand, the Transmission View can also be interpreted as claiming that almost all the explanatory work is done by the IME involved in the explanation. I think this is wrong. The bridge principle connecting the mathematics and the world is far from trivial. We should think hard about what this connection is, and what mathematics contributes. We will return to this point in section 2.5 below.

### 2.4.2 Lange on the Transmission View

In a further important contribution to the debate, Marc Lange (2013) discusses Steiner's proposal, i.e. the Transmission View. Lange is also critical of Steiner's position:

> [N]one of the mathematical explanations in science that I have

> mentioned [including the Königsberg case, TR] incorporates a mathematical explanation in mathematics. These mathematical explanations in science include mathematical facts, of course, but not their proofs – much less proofs that explain why those mathematical facts hold. [...] I am not sure that there even is a distinction between a proof that explains this fact and a proof that merely proves it. (Ibid., pp. 24)

The problem with Lange's position is identical with that of Baker's: he too seems to presuppose that it is necessary for an IME to include a proof. I think this is wrong, as I argued above: It is reasonable to accept Euler's theorem, together with the fact that the Königsberg graph does not satisfy the conditions specified in the theorem, as an explanation of the fact that there is no Euler path on the Königsberg graph. This explanation does not cite the proof of Euler's theorem, it only relies on the fact that it is indeed a theorem.

Can Lange (and Baker) deny that the example of an IME without proof that I just described is in fact an acceptable explanation? They could argue that Euler himself, and mathematicians in general, do not describe their work as involving explanations – this is a problem for both kinds of IMEs, with and without proofs.

A convincing case for IMEs without proofs can be made by comparing the Königsberg IME with its SEM "twin":

1. **SEM** *Explanandum:* there is no path in the bridge system of Königsberg that uses every bridge exactly once, i.e. there is no Euler path in Königsberg. *Explanans:* given a set of reasonable presuppositions, the bridge system of Königsberg has the structure of a graph in which all four vertices have odd valence. According to Euler's theorem, a graph has an Euler path if and only if all, or all but two, vertices have even valence.

2. **IME** *Explanandum:* there is no Euler path on the Königsberg graph. *Explanans:* in the Königsberg graph, all four vertices have odd valence. According to Euler's theorem, a graph has an Euler path if and only if all, or all but two, vertices have even valence.

Both Baker and Lange accept the SEM version as a (good) explanation. Given that this is so, I think it is hard to deny that the IME version is a good explanation as well. I do not see how one can accept one of the above explanations and not the other – this, however, is what they are committed to.

## 2.5 Königsberg in Application

In this section, we shift or focus to the question of how to conceive of the Königsberg case from the perspective of scientific explanations, or SEM.

First, we would like to understand what kind of scientific explanation we are dealing with here. Second, we would also like to obtain an account of how the SEM and IME versions of explanations are related. This, third, could shed light on the connection between pure mathematics and the world; the bridge principle connecting mathematical and empirical structure. We will, again, approach these questions based on Euler's papers and the letter in which the Königsberg Problem was formulated.

### 2.5.1 Presuppositions of the Brute Force Method as an SEM

We distinguished several IMEs in Euler's paper, the Brute Force Method, the Intermediate Method, and Euler's Method, and analyzed them in terms of explanatory power. I proposed that we can give all of these methods both an IME and an SEM reading. I think that these methods do not essentially differ when it comes to their relation to the world. The advantage of the Intermediate Method over the Brute Force Method, and that of the Euler Method over the Intermediate Method, is based on introducing, and exploiting, notation.[12] The innovation is purely mathematical and has nothing to do with their relation to the world. I will therefore limit myself to examining and comparing the IME and SEM version of the Brute Force Method. The hope is that if we understand the relation of the IME and the SEM version of this method, we will automatically have a better understanding of the SEM interpretation of the other explanations.

We begin our analysis with the question of what explicit and implicit assumptions are necessary to formulate the Königsberg Problem as a mathematical problem. What do we assume when we choose the map, and the labels, as the starting point of our mathematical investigation? An analysis, and comparison, of the historical sources suggests that we can distinguish at least three kinds of assumptions.

First, Ehler relied on the map in figure 2.1 to formulate the Königsberg Problem. This provided Euler with a preselection of what counts as an acceptable solution of the Königsberg Problem: Ehler is interested in "structural" answers, i.e. answers related to facts about how the parts of the city hang together. This includes an implicit presupposition that the system of paths is stable over time.

Second, most details about the city of Königsberg are irrelevant and therefore left out. I have already pointed out one example: the information

[12] The importance of notation, in particular in explanatory contexts, has recently been pointed out by Mark Colyvan (2012, ch. 8). It would be worthwhile to explore the ramifications of the present case for Colyvan's ideas.

about the direction of flow of the river is contained in Ehler's map, but omitted in Euler's map. On the other hand, relevant aspects are highlighted in Euler's map by adding labels. By assigning lower-case letters to bridges, the crossing of a bridge is turned into a fundamental process, the details of which are irrelevant. Subsequently, the crossing of several bridges can be written as a sequence of lower-case letters.

Third, both map and labels represent structural constraints on possible paths. We presuppose that we can cross the bridges in certain sequences only, that areas with distinct letters, say $C$ and $D$, are disjunct, and that the map is "complete", i.e that there are no bridges outside the map.

How should we interpret these presuppositions? What do they tell us about the kind of explanation we are dealing with?

### 2.5.2 The Role of Causality

When it comes to scientific explanations, causal explanations are something like the gold standard. Here we will examine whether we can shed light on the Königsberg Problem and the Brute Force Method from a causal point of view.

The Brute Force Method uses strings of lower case letters to denote paths in the bridge system of Könisgberg. We can interpret the lower case letters as standing for causal processes: the label $x$ stands for the crossing of bridge $x$, where we do not care about the exact path that is taken, or the direction in which the bridge is crossed.

The structure of the bridge system then restricts the possibilities of how these causal processes can be combined. The geometry of the situation dictates that not any succession of lower case letters is a path. We operate under the assumption that certain fundamental causal processes can only be combined in a certain way.

The strings of lower case letters thus encode causal-cum-geometrical information. Does this mean that the Brute Force Method provides a causal explanation? Not within the traditional, narrow conception of causal explanations. Traditionally, causal explanations are explanations that specify a cause of a phenomenon, event or fact that is to be explained. The Brute Force Method does not provide a single cause that would explain the phenomenon, the impossibility to travel the system in a certain way.

However, the explanation by the Brute Force Method is not acausal in that it provides relevant causal information. The explanation is of the form: given that certain fundamental causal processes (crossing bridges) can only be combined in a certain way, is it possible to find a sequence such that every fundamental causal process occurs exactly once? The Brute Force Method draws on the causal network of the Königsberg bridge system and is a causal explanation on this wider reading of causal explanations.

One could object that this is not much more than a causal interpretation of a mathematical structure, and a mathematical explanation based on that structure, which is only introduced after the explanatory work is essentially done. Is the causal interpretation of the structure more than the introduction of a trivial bridge principle?

This perspective on the Königsberg Problem, and its solution, is natural only after we have already accepted a certain perspective on that problem. Is it really obvious that we have to answer the question as to why it is impossible to cross all the bridges exactly once based on the map and the structural constraints? No. The context, and the way in which the question is formulated, go a long way towards suggesting what kind of answer we will consider to be acceptable, and the mathematical solution removes all ambiguity. These issues are related to the pragmatic aspects of explanations, to which we will now turn.

### 2.5.3 The Role of Pragmatics

The historical starting point of the Königsberg case is a request for an explanation, as we saw in section 2.5.1. The question was whether it is possible to cross all the bridges in Königsberg exactly once and return to the starting point. We also saw that the question was asked with certain presuppositions. It is fruitful to reconstruct the historical case using Bas van Fraassen's pragmatic theory of explanations, as proposed in van Fraassen (1980, ch. 5).

Van Fraassen's theory is based on the idea that explanations can be interpreted as answers to why questions. Van Fraassen suggests that explanations do not simply cite causes of an event to explain it – '$x$ causes $e$' is not sufficient for '$x$ explains $e$' – the cause has to be salient, and salience is determined by the context in which the explanation is requested. At least two contextual factors determine which causes are salient. The first contextual factor is *relevance*. For example, the person from which we request an explanation helps to determine relevance. Assume we ask a physician about why the victim of a car accident died, then the explanation we expect will have to do with the victim's injuries or the level of alcohol in his blood, but not with the condition of his car or other circumstances of the accident. In this case, the context dictates that the *relevant* causes are medical.

The second contextual factor is the *contrast class*. Using one of van Fraassen's own examples, contrast classes can be illustrated as follows. The question 'Why did Adam eat the apple?' can be interpreted in several ways depending on emphasis. Compare (1) 'Why did *Adam* eat the apple?' with (2)'Why did Adam eat *the apple*?' In (1), we ask why it was Adam, as opposed to some other person or animal, who ate the apple, whereas in (2), we ask why Adam ate the apple as opposed to eating some other fruit, or pastry, or what have you. The difference is one of contrast: as the contrast varies, different kinds of answers or explanations are acceptable. If the contrast is as in (1), the answer 'because he does not like other fruit' is not acceptable, because this was not what we are after.

Both factors help in determining which answer to a why question we find acceptable. Van Fraassen points out that, most of the time, relevance and contrast class will be implicitly given by context, such that all participants of a conversation know what kind of explanation is acceptable. This means that the participants share a set of presuppositions, i.e. assumptions in the background of a request for an explanation.[13]

Let us now locate our historical case in van Fraassen's framework. First, the why question prompting the explanation is: why is it impossible to cross all bridges in Königsberg exactly once and return to the starting point?

Second, relevance is expressed by the maps and the notation. Ehler, who asked the question, was interested in *structural* reasons for the failure of finding a certain path, by providing a sketch of the situation in Königsberg in his letter. Euler further narrowed down the relevant aspects of the structure, by omitting irrelevant details and introducing the letters for areas and bridges. This means that Euler's search for an appropriate mathematical representation was, at the same time, a search for a representation that captures the explanatorily relevant factors, as requested in the formulation, and in the context, of the initial why question. Thus, the transition, from the initial request for an explanation to the reformulation in a mathematical framework, also plays an explanatory role in that the reformulation makes the pragmatic relevance relation explicit.

What, thirdly, is the contrast class? We are asking what distinguishes the situation in Königsberg from other possible bridge systems. What is special about this particular configuration of bridges that stands in the way of an Euler path? If we rely on the fact that relevance is determined by the mathematics, the contrast class of this explanation comprises those bridge systems in which we take the same fundamental causal processes as given, and in which an Euler path exists.

In principle, all three methods discussed by Euler can answer the initial why question, and specify the members of the contrast class. However, in the case of the Brute Force Method, it will be practically infeasible to determine whether a big bridge system is in the contrast class or not. If we take

[13] Van Fraassen's segregation of pragmatic aspects into relevance relation and contrast class is not without complications. To point out just one problem, it is not clear that relevance and contrast class are clearly separate pragmatic aspects: We could think of the contrast class as just one further pointer to the relevant kind of cause or explanans. But then, the contrast class could be integrated into the relevance relation without loss; see Jakob (2007, section 2.3) for criticism along these lines. I think that these problems do not undermine the overall relevance of the pragmatic theory of explanations. What matters is that we can think of explanations as answers to why questions, that we ask these questions with certain implicit presuppositions in mind, which are closely related to the context, and that the context determines which explanations, or answers to the why question, are acceptable. However, see Sandborg (1998) for a critical analysis of van Fraassen's theory, in the context of mathematics.

complexity seriously, then this method will not identify all elements of the contrast class. The Euler method, on the other hand, will give a maximally informative answer to the why question, in that it separates all possible bridge systems into two classes, and specifies a property that all, and only, the members of the contrast class have, namely the even valence of all areas.

This reconstruction does not fit with van Fraassen's theory, in that the relevance relation does not select the explanatorily relevant *causes*, and the contrast class also does not contain contrasting causes. This was to be expected, as we are not dealing with a causal explanation in a narrow sense. However, the *explanans* can be interpreted as giving us structural information about how certain causes hang together, and this is contrasted with other ways in which causes could hang together, and van Fraassen's theory seems to fit nicely with this wider notion of explanation, which is certainly not acausal.

Within this reconstruction, the initial why question is a question about the real bridge system; it is not a purely mathematical question about graphs. The determination of the explanatory relevant factors is at least partially located in the translation of the question about the real system into the question about the graph. The determination of explanatory relevant causal factors becomes obsolete once we conceive of the question as purely mathematical. These two ways of framing the explanation should not be conflated.

## 2.6 Philosophers on Königsberg in Application

In this section, I examine some recent philosophical contributions which examine the Königsberg case as an SEM.

### 2.6.1 Pincock 2007: Abstract Explanation

Christopher Pincock (2007) proposes interpreting the Königsberg Problem as an instance of *abstract explanations*. These are explanations that pick out certain relations of a physical system, while other aspects of the system are ignored. Abstract explanations can rely on mathematics, by using a structure-preserving mapping between the physical system and a mathematical domain, but this mapping does not depend on an arbitrary choice of units or a coordinate system – it captures an intrinsic feature of the system. Here I discuss several different readings of the notion of abstract explanation in Pincock's paper.

On a first reading, the graph-theoretic formulation of the problem amounts to not much more than an incomplete description of the bridge system. Pincock writes: "All that I have done is described the physical system at a higher level of abstraction by ignoring the microphysical properties of the bridges, the banks and the islands" (Pincock, 2007, p. 259). From this perspective, the mathematical description of a physical system is nothing more than a description that picks out certain salient properties of the system, while leaving out others. Or, to put it differently, when we give an abstract description of a situation, we tell the truth, and nothing but the truth, but not the whole truth.[14]

This account of how the explanation works is only partially correct. I argued above that leaving out irrelevant details is an important part of the process, which we can discern in the historical path to the formulation of the explanation. However, it is not all that matters. If my claim, that the mathematical formulation of the problem has a pragmatic aspect, is correct, then the relation between the real bridge system and the mathematical graph potentially violates the "nothing but the truth" clause. For example, it may be possible to find an Euler path in the system if we took alternative paths in the real bridge system into account, but such possibilities are simply excluded for the explanation's sake. The formulation of the problem in graph-theoretic terms excludes a whole class of other reasons as to why it may or may not be possible to travel within the *real* bridge system in a certain way.

Pincock seems to deal with abstraction differently, when he writes that abstract explanations are explanations that "[appeal] primarily to the formal relational features of a physical system" (Ibid., p. 257). This is problematic because Pincock appears to ascribe mathematical, formal properties to the bridge system, which essentially amounts to a mathematization of the real bridge system.

Later on, Pincock retracts this claim. He cautions us against accepting that there are mathematical properties in the world:

> It is tempting to say that the bridge system just is a graph, although this is somewhat misleading. The bridge system is of course not a graph because graphs are mathematical entities and the bridge system is physical. Still, the bridge system and this particular graph seem [...] intimately connected [...]. We might capture this by saying that the bridge system has the structure of a graph, in the sense that the relations among its parts allow us to map those parts directly onto a particular graph. (Pincock, 2007, p. 259f)

According to this quote, we should not think of the bridge system as being identical to a graph, because this implies that there are mathematical properties in the world. Rather, there is a close, intrinsic connection between a particular mathematical structure and the bridge system via a "direct mapping" that preserves graph structure.

[14]The "nothing but the truth" part distinguishes abstraction from idealization: in abstraction, we leave out certain properties in our description of a system, while in idealization, our description of the system comprises claims that are, strictly speaking. wrong; see e.g. Batterman (2010) for a recent discussion.

The idea that the Königsberg graph is connected to the Königsberg system by a structure-preserving mapping is *prima facie* attractive, but proves to be problematic on closer inspection. Two problems discussed in the literature are particularly relevant to the Königsberg case.[15]

The first is the so-called *assumed structure* problem. This is the problem that if we claim that there is a structure-preserving mapping between an empirical and a mathematical domain, we have to assume that the empirical domain is structured into objects and relations such that it makes sense to define a structure-preserving mapping between the two domains. This raises the question as to why this should be so: we simply have to assume that the structure is there to be preserved.

The second relevant problem lies in accounting for *explanatory contributions* of mathematics. If all a mathematical structure does is to represent some structure in the world, how can mathematics be explanatorily helpful? Why don't we base the explanation directly on the structure in the world, which, on this account, is the same as the mathematical structure?

The root of both of these problems is that on the mapping account, we project a mathematical structure, a certain graph, directly onto the world, and claim that this accounts for the contribution of mathematics in the explanation. I think that the reconstruction of the Königsberg case presented above successfully solves both problems. If we take the pragmatic aspect of the explanation into account, then there is no direct mapping between the mathematical structure and the world. The relation between the mathematics and the world is more complicated: the mathematical structure captures our explanatory interest in the system, which need not have the same structure. Furthermore, I argued in section 2.3 that there is an explanatory contribution at the level of pure mathematics, provided that we conceive of the Königsberg Problem as a request for an IME. This explanatory contribution carries over to the bridge system, via the pragmatically-constituted bridge principle.

In sum, neither is the representation relation between mathematics and the world direct, nor is the role of mathematics purely representational, but rather genuinely explanatory at the level of pure mathematics.

### 2.6.2 Pincock 2012: Acausal Representation

In his recent book, Christopher Pincock (2012) classifies the Königsberg case an abstract *acausal* representation – acausal because it does not represent change over time. Pincock writes:

> The mathematics here is not tracking genuine causal relations, but is only reflecting a certain kind of formal structure

[15]See Bueno and Colyvan (2011) for the original formulations of these problems. I review these problems of the so-called *mapping account* in detail in section 7.2.2.

> whose features in the physical system have some scientific significance (Ibid., p. 53)

I agree with Pincock that the graph does not represent change over time. However, I think it is wrong to classify the representation as acausal for this reason. I proposed above that we can give the components of the structure, i.e. the edges, a straightforward causal interpretation. In this sense, the mathematical structure *is* tracking genuine causal relations.

### 2.6.3 Lange 2013: Causality vs. Necessity

Marc Lange (2013) discusses the Königsberg case as an example of so-called *distinctively mathematical explanations.* These are scientific explanations in which mathematics plays a distinctive role. Lange thinks that they do not even fall under a broad notion of causal explanation: "Distinctively mathematical explanations are 'non-causal' because they do not work by supplying information about a given event's causal history or, more broadly, about the world's network of causal relations" (Ibid., p. 3). Rather, such an explanation provides information that is modally stronger than causal information. It works "by showing how the explanandum arises from the framework that any possible causal structure must inhabit, where the 'possible' causal structures extend well beyond those that are logically consistent with all of the actual natural laws there happen to be" (Ibid., p. 21). The explanatory power is derived from the established fact that the explanandum holds as a matter of mathematical necessity, and not from the causal information that it may provide.

Lange grants that we can give the parts of the Königsberg graph a causal interpretation, and that we presuppose causal stability of the bridge system for the explanation. Distinctively mathematical explanations can cite causes, or rely on causal structure. This, however, does not undermine the fact that this is a case of a distinctively mathematical explanation:

> [T]he fixity of the arrangement of bridges and islands, for example, is presupposed by the why question that the explanation answers: why did this attempt (or every attempt) to cross this particular arrangement of bridges – the bridges of Königsberg in 1735 – end in failure? [...] [T]he why question itself takes the arrangement as remaining unchanged over the course of any eligible attempt. If, during an attempt, one of the bridges collapsed before it had been crossed, then that journey would simply be disqualified from counting as having crossed the intended arrangement of bridges. The laws giving the conditions under which the bridges' arrangement would change thus do not figure in the explanans. (Likewise, it is understood in the why

> question's context that the relevant sort of 'crossing' involves a continuous path.) (Ibid., p. 13)

Here Lange claims that the explanatory work is done by the mathematics, while all the causal presuppositions (laws) are fixed by the why question, and therefore do not figure in the explanans. However, later on, he appears to claim that at least some contingent, causal facts are part of the explanans:

> The explanans consists not only of various mathematically necessary facts, but also [...] of various contingent facts presupposed by the why question that the explanandum answers, such as that the arrangement of bridges and islands is fixed. The distinctively mathematical explanation shows it to be necessary (in a way that no particular force law is) that, under these contingent conditions, the bridges are not crossed. (Ibid., pp. 21)

Lange's account of the Königsberg case shows some similarities with my account; however, several aspects can be clarified by rephrasing them in my framework.

One similarity is that Lange's account is compatible with a pragmatic account of the Königsberg case. He agrees that the why question, and the context, help in determining a subset of acceptable answers, and in the last quote, he also seems to endorse the view that some contingent, causal assumptions are part of the explanans. However, if this is correct, it is unclear to me how he can maintain that the explanatory power has to be located exclusively in the mathematical demonstration, or that the explanation is non-causal, even on the broad conception of causal explanation. This is impossible, if causal information is part of the explanans. Here Lange is ambiguous.

In my account, mathematics contributes in two ways to the SEM version of the Königsberg Problem. First, the mathematics is explanatory at the level of pure mathematics by accounting for the purely mathematical fact that no Euler path exists in the Königsberg graph. Second, the mathematics has an explanatory role, in the transition from the mathematical structure to the real bridge system, by making explicit what the causal factors of interest are, in this request for an explanation.

We have seen in section 2.4.2 that Lange is opposed to the notion of IME, because he thinks that no clear notion of explanatory proof is to be had. I, on the other hand, argued that IMEs need not be proofs, but can be applications of theorems to mathematical facts. If we accept this, it is no wonder that Euler's explanation is a matter of mathematical necessity: we can reconstruct it as an *intra*-mathematical explanation, a mathematical deduction. I think that Lange confuses the IME version with the SEM version by projecting the mathematical structure directly into the world.

The application of Euler's IME to the world is a matter of pragmatics, and Lange would probably concede that much. I would argue, further, that this is where causal information comes into play, turning this into a causal explanation in a broad sense. The connection between explanans and explanandum is not a matter of mathematical necessity, as causal presuppositions feature in the explanans. In this step, the mathematics plays an explanatory role, by capturing the salient causal structure, and by restricting the class of acceptable answers to the why question.

Euler's solution to the Königsberg Problem is so successful that it leads to a gestalt switch. Once we have seen his solution, and the precise mathematical formulation of the question that is possible in the framework of graph theory, we forget that this framework, and the question formulated in it, both come with certain presuppositions that are pragmatic in nature. I think that this is what happened to Lange: the why question, and the context in which it is asked, is already heavily mathematized, as witnessed by the historical sources.

## 2.7 Conclusions

Here are the main question, lessons, and open problems of this chapter.

First, a methodological remark. Once more, the close examination of the original scientific exposition of a problem, and its solution, prove to be very fruitful. We can track the genesis of the formulation of the two problems, and we discover that Euler proposed several solutions, including a discussion of their respective advantages and drawbacks. This settles, once and for all, the notion that not all solutions to one and the same mathematical problem are equal; additionally, the interpretation of the differences, in terms of explanatory power, seems to be instructive.

Second, we discerned two notions of IMEs, as instantiated by the solutions to the General Problem and the Königsberg Problem. The former involves a proof of a theorem, or a method for solving a decision problem, while the latter consists of the application of that theorem, or method, to a particular instance. This distinction helped us clarify issues in the philosophical debate; in particular, it calls into question arguments by Alan Baker and Marc Lange against the Transmission View. The Transmission View stands as a viable option, with the caveat that the bridge principle can involve more than mere structure-preservation. The distinction between these two kinds of IMEs need not be exhaustive. There are probably further kinds of IME that we have not touched upon.

Third, there are at least three different methods for solving the General Problem in Euler's paper, the Brute Force Method, the Intermediate Method, and Euler's Method. Euler himself discusses the systematic differences between these methods. We interpreted these differences in terms of explanatory power. The ideas, that (computational) complexity, along with the provision of just the right amount of information, are maximized in the form of an equivalent characterizing property, seem particularly promising. These ideas are not fully worked out yet, and should be further explored.

Fourth, the invention of graph theory is closely connected to a notational innovation: the idea to write some components of the Königsberg system, the bridges, in terms of another, the areas connecting the bridges. This, in turn, is the basis for the reduction of complexity, and enhanced explanatory power obtained in the transition from the Brute Force Method to Euler's Method. The relation between these issues, and a potential explanatory role of notation, is also an issue that should be investigated further.

Fifth, interpreting the Königsberg Problem as a request for an SEM raised the question as to what kind of scientific explanation this is. The proposed solution has two aspects. First, this is not a causal explanation in the narrow sense that it cites a cause of the explanandum. However, if we adopt a wider notion of causal explanation, which requires that the explanans relies on causal structure, then this is a causal explanation, given that the components of the mathematical structure have a causal interpretation. Second, the formulation of the explanation in graph-theoretic form is pragmatic, in van Fraassen's sense, in that the explanation excludes whole classes of causes that are potentially explanatorily relevant, and focuses on the "structural" reasons as to why it is impossible to cross all the bridges exactly once.

Sixth, I argued against two accounts of the Königsberg case as a kind of scientific explanation. Christopher Pincock's notion of abstract explanation was found to be only partially adequate in that it does not mesh with the pragmatic aspect of the explanation, which implies that the explanation does not only rely on the omission of certain details of the system, but rather on distorting the causal structure for the explanation's sake. Marc Lange's account of the case as a distinctively mathematical explanation loses its plausibility, if we keep in mind that we can reconstruct the solutions to the Königsberg problem both as SEM and IME.

To prevent the danger of overgeneralizing the systematic significance of these results, the next step in the investigation would be, firstly, to push the systematic questions just mentioned further, and, secondly, to apply the concepts, and theses, developed here to other, ideally more complicated, examples of explanations involving mathematics.

# Chapter 3

# The Lotka-Volterra Predator-Prey Model

## 3.1 Introduction

The topic of this chapter is the predator-prey model from population ecology, proposed by Lotka and Volterra in the beginning of the 20th century.[1] The Lotka-Volterra predator-prey model, predator-prey model for short, is a case of mathematical modeling in biology that has received much attention in philosophical debates, and was a seminal early contribution to theoretical population ecology. I will revisit the historical papers by Vito Volterra that introduced the model; the focus will be on the role of mathematics in application to population systems. More specifically, I will closely examine the empirical puzzle that was responsible – or so we are told – for the construction of the model.

After following the historical discussion of the model in time and giving a brief account of the model's status in population ecology today, I extract some interesting philosophical lessons from Volterra's original account of the model. I then relate the historical account and the philosophical lessons to some recent philosophical work that draws on the predator-prey model as a case study.

I discuss the following philosophical issues and questions.

- **Idealization and Mathematical Modeling:** What is the historical and systematic status of the idealizations of the model; how are the idealizations justified? Are the idealizations driven or suggested by the mathematics or by empirical considerations? What is the historical

[1] This chapter draws on my paper Räz (2013a), as well as Scholl and Räz (2013). Some passages of the chapter report on an ongoing project with Raphael Scholl. I thank Raphael for many discussions about this episode. I am also very grateful to Claus Beisbart for extensive comments on a previous draft of this chapter.

and systematic perspective of scientists and philosophers on the use of mathematics in biological models?

- **Causality and Mathematical Modeling:** What is the role of causality in this model? What is the relation between causal inference and modeling as scientific practices? Can we interpret parts of the model causally? What is the role of intervention in this model?

- **Robustness Analysis:** What are the ramifications of a historical perspective on the case for robustness analysis? Is robustness analysis a topic that is only of interest to population ecology? What is the relation between the precise mathematical formulation of properties of the model and the robustness of these properties?

### 3.1.1 Overview

In section 3.2, I present Volterra's 1928 account of the predator-prey model, and the deduction of the so-called Third Law, which is supposed to explain certain features of fishery statistics. I follow the mathematical derivations in some detail.

In section 3.3, we have a look at the immediate reactions of population ecologists to Volterra's proposal, again with a focus on the role of mathematics. We will see that at the time, theorists took some of the idealizations to be particularly worrisome.

In section 3.4, I discuss some interesting passages from Volterra's and d'Ancona's later writings, which can be interpreted as methodological reflections that were prompted, at least partially, by the early critics of the model. Volterra and d'Ancona give an extended motivation for using mathematical modeling instead of some other mode of theorizing.

In section 3.5, I recapitulate the status of the predator-prey model in population ecology today. I reconstruct the reason why the model is commonly not accepted as providing an account of actual predator-prey interactions. The model is nowadays taken to serve pedagogical purposes, and it is used as a template for more complicated, and realistic, models of predator-prey systems.

In section 3.6, we begin to harvest the philosophical fruit of our previous work. I discuss: Volterra's remarks on mathematical methodology; the relation between mathematics and idealization in the construction of the model; some relevant details in the derivation of the Third Law; and notions of intervention that are motivated by one of Volterra's main mathematical tools, phase spaces.

In section 3.7, I put the preceding philosophical analysis, and the detailed reconstruction, to work in some recent philosophical debates on the predator-prey model. In section 3.7.1, I comment on a recent paper by Mark Colyvan (2013); I argue that his main thesis, that mathematical models in the special sciences can be explanatory, is supported by the historical sources. I further point out that Colyvan underestimates the importance of idealizations in the model. In section 3.7.2, I critically examine the recent contribution of Weisberg and Reisman (2008) to the debate on robustness analysis. I claim that Weisberg's and Reisman's account suffers from mathematical imprecisions, which threaten to undermine their main thesis that Volterra's Third Law is a robust property of a whole class of predator-prey models. I also cast doubt on robustness analysis as a phenomenon of biology, as opposed to other modeling sciences. Section 3.7.3 has some critical remarks on Christopher Pincock (2012); I argue that his view, that the predator-prey model is an acausal representation, is mistaken.

Section 3.8 sums up the main points of the chapter, and notes open questions.

## 3.2 The Predator-Prey Model in 1928

In this section, I present the predator-prey model, based on the first historical in-depth account by Vito Volterra (1928).[2] According to Volterra, the construction of the model was prompted by a request for an explanation; see Volterra (1928, p. 4, fn 2). Umberto d'Ancona, a marine biologist and Volterra's son-in-law, brought the following puzzle to Volterra's attention. Fishery statistics showed an increase in the number of predators relative to the number of prey in the adriatic sea during the first world war, a period in which fishing diminished; the pre-war proportion was restored when fishing returned to its old intensity.

Volterra was able to qualitatively reproduce and explain this surprising proportion shift with the predator-prey model. The central result for the explanation of the phenomenon is Volterra's *Third Law*, the "Law of the disturbance of the averages". In the 1928 paper, Volterra had already shifted his focus from the mere explanation of the proportion shift to the goal of a quantitative theory of interspecies relations, taking into account interspecies competition, interaction between $n$ species, and more complex interaction patterns. Here I will focus on the original predator-prey model, the derivation of Volterra's Third Law, and the explanation of the proportion shift.

### 3.2.1 Volterra on the Role of Mathematics

In the beginning of the paper, Volterra reflects on the use of mathematical methods in population biology. He first addresses general worries about the adequacy of mathematical methods in application to complex biological systems:

[2]The 1928 account is predated by Volterra (1926), which is much shorter, and Alfred Lotka's writings on the same model, which I will not discuss here.

> [O]n first appearance it would seem as though on account of its extreme complexity the question might not lend itself to a mathematical treatment, and that on the contrary mathematical methods, being too delicate, might emphasize some peculiarities and obscure some essentials of the question. To guard against this danger we must start from hypotheses, even though they be rough and simple, and give some scheme for the phenomenon. (Ibid., p. 5)

In the discussion of the modern perspective on the Lotka-Volterra model in section 3.5, we will see that his concer about "peculiarities", generated by mathematics, was justified. However, the problem is not due to the use of mathematical methods in general, but rather due to problems of this particular model. While the justification of the use of mathematics here seems all too brief, Volterra's defense of mathematical methods became more sophisticated in later publications, as we will see in section 3.4.

Having addressed some of the problems that the use of mathematical methods might cause, Volterra turns to a positive characterization of the role of mathematics in modeling:

> And what mathematical methods will it be convenient to use? [...] Permit me to indicate how the question can be considered: Let us seek to express in words the way the phenomenon proceeds roughly: afterwards let us translate these words into mathematical language. This leads to the formulation of differential equations. If then we allow ourselves to be guided by the methods of analysis we are led much farther than the language and ordinary reasoning would be able to carry us and can formulate precise mathematical laws. These do not contradict the results of observation. Rather the most important of these seems in perfect accord with the statistical results [...]. The road followed is thus clearly indicated with these few words. We shall see after a little how the difficulties met were overcome. (Ibid., pp. 5)

Volterra distinguishes at least four steps of the application process. First, we start with an informal, but abstract description of a phenomenon of interest. Second, we translate the informal description into mathematical language, which results in differential equations. Third, we deduce laws from the equations, using results from qualitative analysis of ordinary differential equations. Volterra thinks that mathematical methods are most fruitful in the third step, as the mathematics facilitates formal inferences, carrying us further than informal ones. In a fourth step, the "mathematical laws" deduced from the model are compared with empirical data. In a footnote, Volterra gives an example for such a "mathematical law"; the Third Law. We will discuss its derivation in detail in section 3.2.4 below.

### 3.2.2 Introduction and Justification of the Model

The core of the Lotka-Volterra predator-prey model is a set of two coupled, non-linear differential equations. Volterra writes them as[3]

$$\frac{dN_1}{dt} = (\epsilon_1 - \gamma_1 N_2)N_1 \tag{3.1}$$

$$\frac{dN_2}{dt} = (-\epsilon_2 + \gamma_2 N_1)N_2 \tag{3.2}$$

These two equations are supposed to describe the evolution of the predator and prey populations over time. The components of the equation are defined as follows:

- $N_1$ is the number of prey.
- $N_2$ is the number of predators.
- $\epsilon_1$ is the net growth rate of the prey.
- $\epsilon_2$ is the net rate of decrease of the predators.
- $\gamma_1$ is the "aptitude of defense" of the prey.
- $\gamma_2$ is the "means of offense" of predators.

All the variables and parameters are assumed to range over positive real values. Volterra devotes quite some space to the introduction and justification of the equations. We can discern two aspects of this discussion.

The first aspect is that the model is not supposed to capture changes in populations that are due to external factors, but rather to capture *purely internal phenomena*. These are phenomena that do not depend on interactions of the fish with the environment, such as migration, or the change of seasons, or impacts of any further species, but only on the interaction of the two species with each other, "due only to the reproductive power and to the voracity of the species as if they were alone" (Ibid., p. 5). If such external factors play a significant role in population dynamics, then the model does not mirror the system adequately.

The second aspect is the justification of the particular form of equations 3.1 - 3.2. It is necessary to introduce idealizing assumptions in order to formulate the model: According to Volterra, we have to assume that the populations are described by continuous, not discrete variables – otherwise we could not use differential equations. We also assume constant, continuous birth and death rates (yielding the net growth), and homogeneity of the

[3] I will use Volterra's historical notation throughout.

individuals – for example, age structure, and the dependence of fertility on age, are neglected.[4]

These assumptions justify the use of constant growth coefficients $\epsilon_1$ and $\epsilon_2$ in the linear terms of the predator-prey equations. These terms determine the time evolution of the two species independent of interactions: in the absence of predators, the prey population grows exponentially, and the predators die out exponentially, as $e^{-t}$.

What remains to be justified, then, are the two non-linear interaction terms. Here is Volterra's story:

> [If] the second species feeds upon the first $\epsilon_1$ will diminish and $-\epsilon_2$ will increase, and evidently the more numerous the individuals of the second species become the more $\epsilon_1$ will diminish, and the more the individuals of the first species increase, the more will $-\epsilon_2$ increase. To represent this fact in the simplest manner let us suppose that $\epsilon_1$ diminishes proportionally to $N_2$, that is by the amount $\gamma_1 N_2$, and that $-\epsilon_2$ increases proportionally to $N_1$, that is by the amount $\gamma_2 N_1$. (Ibid., p. 9)

Volterra believes that the interaction terms modify the natural growth and death rates of the two species in the "simplest manner", viz. as proportional to the size of both species. However, he is not entirely satisfied with this argument; he adds two further justifications. First, he proposes to take the *probable number of encounters* into account to support the form of the interaction terms. Second, he claims that the results he deduces later are valid for more general interaction terms.

The justification of the interaction terms based on the probable number of encounters is given in §4. It runs as follows. The number of encounters between the two species in one unit of time can be assumed to be proportional to $N_1 N_2$, i.e. equal to $\alpha N_1 N_2$. Then, the two species will be affected differently by the encounters. In the case of one species preying on the other, the predators will profit by a positive factor $\beta_2$, while the prey will experience a different, negative factor $\beta_1$. We can now essentially set $\beta_1 \alpha = -\gamma_1$ and $\beta_2 \alpha = \gamma_2$, which yields the interaction terms of the predator-prey model. Volterra does not elaborate further on the second justification; the deduction with general interaction terms.

### 3.2.3 Exploring the Model

After setting up the model, Volterra turns to its mathematical analysis, in particular the derivation of his three laws. First, he notes that the quotients

[4] It is not clear that all these assumptions are necessary. The parameters could also be interpreted as averages, such that, say, homogeneous individuals are not a necessary assumption.

$K_2 := \frac{\epsilon_1}{\gamma_1}$ and $K_1 := \frac{\epsilon_2}{\gamma_2}$ are the components of the stationary state of the system; i.e. if you insert these numbers into the predator-prey equations, they are equal to 0, meaning that if population sizes are equal to these quotients, population sizes do not change over time.

Next, he integrates the equations: First, he removes dependent parameters – this is called non-dimensionalization, see Murray (1993, p. 64) – by setting $n_1 := \frac{N_1}{K_1}$ and $n_2 := \frac{N_2}{K_2}$. This yields the equations

$$\frac{dn_1}{dt} = \epsilon_1(1-n_2)n_1 \tag{3.3}$$

$$\frac{dn_2}{dt} = -\epsilon_2(1-n_1)n_2 \tag{3.4}$$

These are then added and integrated to yield

$$\left(\frac{n_1}{e^{n_1}}\right)^{\epsilon_2} = \mathbf{C}\left(\frac{n_2}{e^{n_2}}\right)^{-\epsilon_1} \tag{3.5}$$

This what is called a first integral, that is, there is a constant relation **C** between the two variables $n_1$ and $n_2$. This is a necessary condition for the subsequent phase space analysis.

Volterra examines how the right hand side and the left hand side of equation 3.5 behave as functions of $n_1$ and $n_2$; then he combines the two functions in phase space. From this, and equation 3.5, it can be deduced that the population sizes $N_1$ and $N_2$ fluctuate in closed, periodic cycles. Finally, he derives that the period is constant. The result is summed up as follows: "We have then in this case a *periodic fluctuation of the number of individuals of the two species, with period T*, or the phenomenon will have a *cyclically periodic* character" (Volterra, 1928, p. 15, emphasis in original).

The population dynamics of the Lotka-Volterra equations can be illustrated using a phase space diagram; see figure 3.1. $\Omega$ is the stationary state of the system. The orbits $X$, $\Psi$, $\Lambda$, $\Phi$ are examples of the different, exhaustive and disjunct orbits describing the evolution of the two population sizes; the evolution runs conter-clockwise. Each orbit corresponds to a separate class of initial conditions, and if the system is not disturbed, it will stay on its orbit. Note that the parameters of the system, i.e. the $\epsilon$s and $\gamma$s, are assumed to be constant.

### 3.2.4 Derivation of the Third Law

In section 7, Volterra turns to the derivation of the Third Law. He first determines the averages of the two populations during a period, and, secondly, how these averages are affected by changes in the $\epsilon$ parameters.[5]

[5] The presentation in this section draws on Braun (1993, p. 447), who has a conspicuous reconstruction of Volterra's derivation.

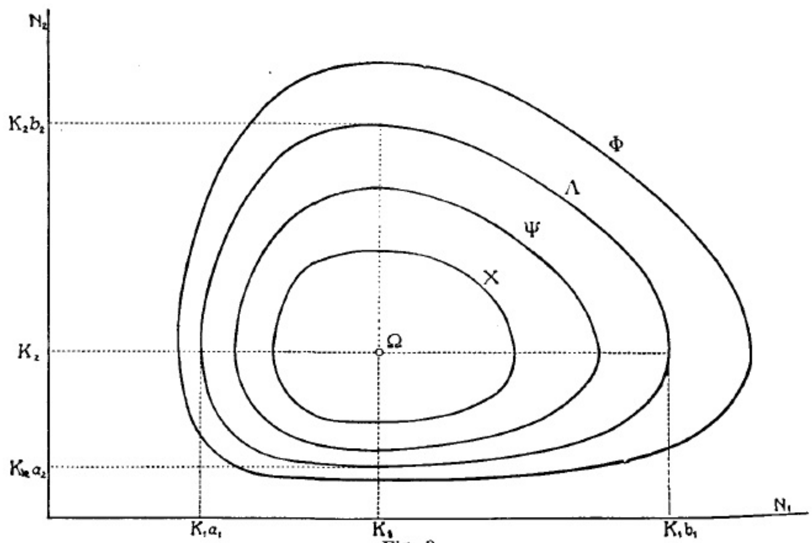



Figure 3.1: Periodic Orbits, from Volterra (1928, p. 14)

First, to the average populations numbers. The time average of a population $N$, $\bar{N}$, over a period $T$, is defined as

$$\bar{N} = \frac{1}{T}\int_0^T N dt \tag{3.6}$$

This average can be determined directly from the predator-prey equations for both species. By dividing both sides of equation 3.1 by $N_1$ we get

$$\begin{aligned} \frac{dN_1}{N_1 dt} &= (\epsilon_1 - \gamma_1 N_2) & (3.7) \\ \frac{1}{T}\int_0^T \frac{dN_1}{N_1} &= \frac{1}{T}\int_0^T (\epsilon_1 - \gamma_1 N_2) dt & (3.8) \end{aligned}$$

Now, because

$$\begin{aligned} \int_0^T \frac{dN_1}{N_1} &= \int_0^T dln[N_1] \\ &= ln[N_1(T)] - ln[N_1(0)] & (3.9) \end{aligned}$$

the left hand side of equation (3.8) is 0; $N_1$ is $T$-periodic, i.e. $N_1(T) = N_1(0)$. From (3.8) we can thus deduce

$$\frac{1}{T}\int_0^T (\epsilon_1 - \gamma_1 N_2)dt = 0 \tag{3.10}$$
$$\frac{1}{T}\int_0^T N_2 dt = \frac{1}{T}\int_0^T \frac{\epsilon_1}{\gamma_1}dt = \frac{\epsilon_1}{\gamma_1} \tag{3.11}$$

The derivation for the second case is analogous. We thus get the result that the average of the population sizes of the two species over a period is identical to the equilibrium point $\Omega = (K_1, K_2)$

$$\frac{1}{T}\int_0^T N_1 dt = \frac{\epsilon_2}{\gamma_2} \tag{3.12}$$
$$\frac{1}{T}\int_0^T N_2 dt = \frac{\epsilon_1}{\gamma_1} \tag{3.13}$$

The most important consequence of this result is that

> [I]f $\epsilon_1$, $\epsilon_2$, $\gamma_1$, $\gamma_2$ stay constant, the averages of the individuals of the two species during a cycle of fluctuation will always be the same whatever may be the initial numbers of individuals of the two species. (Volterra, 1928, p. 18)

Thus, on whatever orbit the system is in phase space, see figure 3.1, the average number of species for one full cycle is identical to the abundance of the stationary state $\Omega$.

Finally, Volterra turns to the Third Law, which concerns the *disturbance* of averages. The question here is how the system is affected if the $\epsilon$-parameters change, i.e. if we interfere with the reproductive abilities of the two species, while the $\gamma$-parameters are held constant.

We modify the original equations by adding a constant $\alpha$, positive or negative, to the natural growth rate of the prey population, and the same constant $\alpha$ to the natural depreciation rate of the prey population

$$\frac{dN_1}{dt} = (\epsilon_1 - \alpha - \gamma_1 N_2)N_1 \tag{3.14}$$
$$\frac{dN_2}{dt} = (-\epsilon_2 - \alpha + \gamma_2 N_1)N_2 \tag{3.15}$$

We assume that $\alpha$ does not change the sign of the overall growth rate of both species, i.e. we assume it to be smaller than both $\epsilon$ parameters. Also, we could have chosen separate $\alpha$s for the two species, but our main interest is when the change in growth rates is the same for both species.

The new system is qualitatively identical to the original system; this can be seen by letting $\epsilon_1^* := \epsilon_1 - \alpha$ and $\epsilon_2^* := \epsilon_2 + \alpha$. We get a new equilibrium point for the new system:

$$\frac{\epsilon_1^*}{\gamma_1} = \frac{\epsilon_1 - \alpha}{\gamma_1} \tag{3.16}$$

$$\frac{\epsilon_2^*}{\gamma_2} = \frac{\epsilon_2 + \alpha}{\gamma_2} \tag{3.17}$$

This equilibrium point is, by the same argument as before, the time average of the number of both species for the new system:

$$\frac{1}{T}\int_0^T N_1 dt = \frac{\epsilon_2 + \alpha}{\gamma_2} \tag{3.18}$$

$$\frac{1}{T}\int_0^T N_2 dt = \frac{\epsilon_1 - \alpha}{\gamma_1} \tag{3.19}$$

Let us now see what happens to the averages if we vary $\alpha$. Assume $\alpha$ is positive. With the help of equation (3.14), we can interpret this as a reduction of the growth rate of the prey population, $\epsilon_1$, by a factor $\alpha$, and from equation (3.15) we see that it means we enlarge the depreciation rate of the predator population, $\epsilon_2$, by the same factor $\alpha$. Put differently, we reduce the growth rates of both populations by a constant rate. The effect of these changes on the time averages is that the average number of prey is increased, see equation (3.18), and the average number of predators is decreased, see equation (3.19). This is Volterra's Third Law.

The scope of the Third Law is quite general. In a footnote, Volterra writes that " this law is valid within certain limits [...] as long as the coefficient of increase $\epsilon_1$ remains positive" (Volterra, 1928, p. 20). The principle is valid as long as the change in parameter $\alpha$ will not result in a change of sign of the $\epsilon$ parameters, which would constitute a qualitative change of the system.

It is, again, useful to examine how the change in parameters affects phase space. Here is Volterra's description:

> In figure [3.2, see below] we have represented the transition from a cycle $\Lambda$ corresponding to the parameters $\epsilon_1$ and $\epsilon_2$ to a cycle $\Lambda'$ corresponding to parameters [$\epsilon_1 - \alpha \leq \epsilon_1, \epsilon_2 + \alpha \geq \epsilon_2$ ...]. We may conceive of this transition as occurring in an instant corresponding to the point P of intersection of the two cycles, that is to say, without having any sensible change at that instant in the number of the individuals of the two species, although a change is disclosed with the passage of time by virtue of the constant action due to the variation of the parameters $\epsilon_1$ and $\epsilon_2$.

> The centre $\Omega'$ of $\Lambda'$ is moved to the right of and below $\Omega$ which indicates a diminution in the average value of $N_2$ and an increase in the average value of $N_1$. (Ibid., pp. 18)

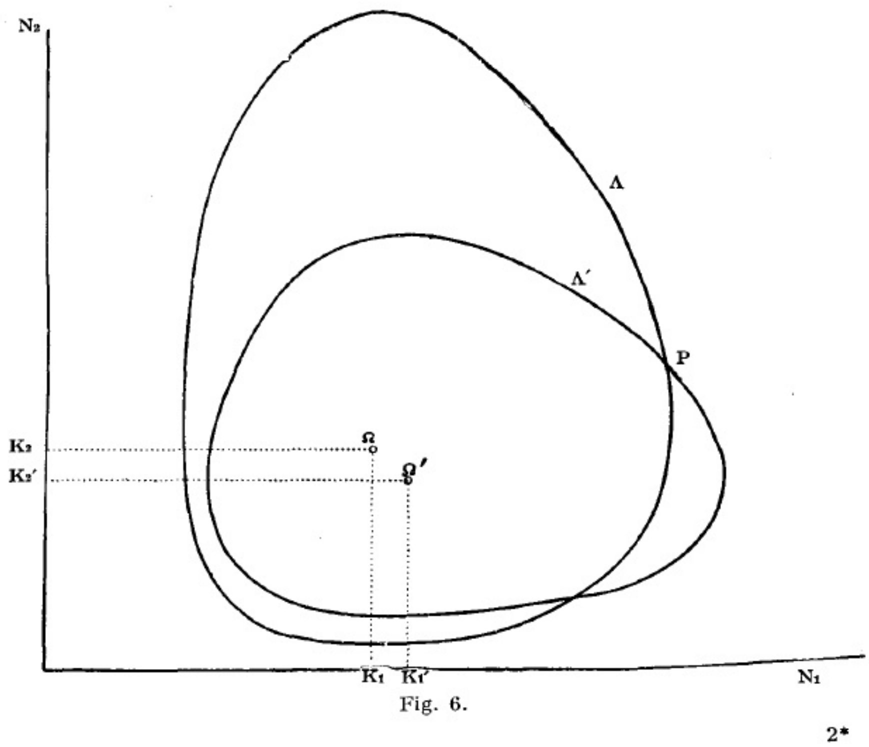



Figure 3.2: Volterra's Third Law, from Volterra (1928, p. 19)

Note that the new trajectory of the system depends on the point at which the parameters are reset: the system would be on a different orbit than $\Lambda'$ if the parameters were reset at the point where $N_2$ is maximal, even if the parameters were the same as those of $\Lambda'$; same phase space, different trajectory.

So far, Volterra's reasoning does not depend, to a significant degree, on the interpretation of the variables and parameters at play, and it is not yet clear how changes in parameters can be implemented. Volterra writes that "constant action" is necessary to "enforce" the new trajectory $\Lambda'$. The change of parameters can be implemented as follows:

> Now to make $\epsilon_2$ increase means destroying uniformly individuals of the second species in a quantity proportional to their number, and to make $\epsilon_1$ decrease means destroying uniformly in-

> dividuals of the first species in a quantity proportional to their number ... (Ibid., p. 18)

And a little bit later, he points out that fishing is an instance of such a uniform destruction:

> It seems that the animal species for which in their natural state the verifications of these laws can most easily be carried out are fish, of which there are in fact species which feed upon others. Continual fishing constitutes a uniform destruction of individuals of the various species. The cessation of fishing during the period of the recent war and its resumption after the war established transitions comparable to those considered above, from one cycle to another. Besides, the greater or less abundance of fish of various species determined by statistics gives a measure of the abundance of the individuals of the various species; hence the statistics of fishery furnish data on the fluctuations. The results of the statistics are seen to be in accord with the mathematical predictions [...]. (Ibid., p. 21)

The mathematical shift of averages, which can be implemented by a continuous, uniform intervention in a two-species system, explains the fishery statistics, which showed the surprising shift in proportions. On Volterra's view, this constitutes a confirmation of the mathematically derived results.

This concludes our reconstruction of Volterra's account of the predator-prey model and the derivation of the Third Law. We will turn to a philosophical analysis of Volterra's work in section 3.6; beforehand, we will follow the history of the model up to its status in contemporary population ecology.

## 3.3 Contemporary Voices on Lotka-Volterra

In this section, we have a brief look at the immediate historical reactions to Volterra's model, with special attention to criticism, be it against the use of mathematical models in general, against particular modeling assumptions, or biological concerns.

Two of the first critics of Volterra's and d'Ancona's work were Egon Pearson and Friedrich Bodenheimer. Their criticism was mainly directed at the applicability of the model to the Adriatic. Pearson worried that:

> [P]erhaps other factors which d'Ancona had not considered, such as changes in the methods of fishing or even migration of the fish, might account for the observations during the war years ... . (Kingsland, 1985, p. 131)

Bodenheimer advanced similar worries about uncontrolled environmental factors. For both critics, the point of contention was not the use of mathematical modeling in population ecology in general, but its applicability to this particular system, the Adriatic. They called into question whether the *isolation* assumption was justified in the case of the adriatic: even if the model reproduces phenomena, such as oscillating populations and the Third Law, it is possible that, in the case of the Adriatic, external causes are responsible for these phenomena.

There is a partial remedy for this problem, namely trying to isolate a different system in an experimental setting. Such an experiment was carried out by Georgii Frantsevich Gause:

> To test the predator-prey model, Gause set up populations of protozoans in test-tube environments [...]. The Lotka-Volterra equations had predicted that the two populations would oscillate continually, without either species going extinct. Gause's populations were not so obliging: his predators quickly consumed all the prey and then died off shortly afterwards. (Ibid., p. 150)

It proved difficult to produce the behavior predicted by the model in an experimental setup that controlled for environmental factors. However, Gause was able to produce oscillations under very specific conditions. First, oscillations occurred when he allowed for "immigration", i.e. adding a number of predators and/or prey from outside the system. A more promising, second finding was that if predation was at a "low intensity", oscillatory behavior did occur. The first result suggests that periodicity could be due to external factors, as proposed by Pearson, while the second can be interpreted as imposing restrictions on the range of parameters (means of offense and defense, $\gamma$ parameters) in which the model is valid.

Apart from specific criticism of the predator-prey model, and attempts to verify the model experimentally, the reactions to the use of sophisticated mathematical methods in population ecology were mixed, ranging from wholehearted endorsement to principled rejection. On one end of the spectrum, Royal Chapman was convinced of the usefulness of mathematical methods in ecology:

> Mathematics, he believed, would raise the lowly status of ecology to the dignified level of the physical sciences. He believed that Volterra's publications were destined to be as important for population biology as those of Willard Gibbs for physical chemistry [...]. Ecology was on the way to being an exact science. (Ibid., p. 128)

On the other end of the spectrum, the prominent entomologist William Robin Thompson was critical of the usefulness of mathematical methods in population ecology beyond heuristics:

> What alarmed him especially was not the continued growth of mathematical ecology per se, but rather his feeling that people had stopped discriminating between mathematical figments and biological facts. [...] Mathematical ecology was no longer just a stimulus to the imagination, a way to help biologists envisage a problem. People were beginning to believe that it held the truth. (Ibid., p. 139)

In sum, these voices represent the whole spectrum of attitudes towards the use of mathematical methods in biology. In hindsight, critics addressing particular modeling assumptions, such as the question of how to de-isolate the system, were right: the isolation assumption is indeed problematic. It is more difficult to evaluate the more sweeping endorsements or rejections of mathematical methods in population ecology. Today the field thrives, and at least some of its exponents have a nuanced view of the usefulness and pitfalls of mathematical models, as we will see in section 3.5 below.

## 3.4 Volterra and d'Ancona 1935: Methodological Reflections

In this section, we revisit Volterra and D'Ancona (1935, Ch. 1), a monograph on population ecology, with special attention to the methodological motivations driving Volterra's and d'Ancona's research.[6] Volterra and d'Ancona reflect on the reasons for adopting mathematical modeling as their method of choice. In a nutshell, Volterra and d'Ancona would have preferred a different, more direct approach to population ecology, but were forced to adopt the modeling path by limited epistemic access to the system under scrutiny.

Their first, preferred method would have been direct causal inference by controlled experimentation. This, however, is not feasible because actual populations cannot be controlled adequately: they are spread out spatially, their breeding cycles are too long, and environmental factors vary indefinitely. All this prevents a direct, experimental approach to *this* system. Second, detailed statistics could compensate for some of the epistemic limitations of controlled experimentation, as varying factors would eventually cancel out over time. However, a statistical approach is out of the question too, because the necessary practical means to carry out statistical evaluations were not available at the time.

As the methods of experimental control and statistics cannot be applied, Volterra and d'Ancona propose a third, more indirect approach. They write:

> Since it appears too difficult to carry through quantitative studies by experiments and thus to obtain the laws that regu-

[6]The following discussion of Volterra's and d'Ancona's methodological reflections is taken from Räz (2013a), which is partially based on Scholl and Räz (2013, Sec. 3).

> late interspecific relationships, one could try to discover these same laws by means of deduction, and to see afterwards whether they entail results that are applicable to the cases presented by observation or experiment.[7]

Volterra and d'Ancona advocate an indirect, modeling approach. However, they only resort to mathematical modeling *faute de mieux*: a more direct approach, based on controlled experiments, is simply too difficult to carry out. The use of mathematics is not due to the desire to apply complicated mathematics at all cost, but due to a lack of alternatives. This means that in order to carry out their explanatory project, the use of modeling techniques was the only viable option. Population ecology faced real methodological difficulties that could only be overcome with the help of mathematical methods.

Volterra and d'Ancona are outspoken defendants of the use of mathematical methods in population ecology that parallel physics. Volterra held a Chair of Mathematical Physics at the University of Rome, and he does not hide his sympathy for mathematical methods:

> One should not worry too much when one considers ideal elements and imagines ideal conditions that are not completely natural. This is a necessity, and it is sufficient to think of the applications of mathematics to mechanics and physics that have led to results that are important and useful in practice. In rational mechanics and in mathematical physics one considers surfaces without friction, absolutely flexible and unextended strings, ideal gases, and so on. The example of these sciences is a great example we should always keep in mind and that we should strive after.[8]

Two aspects of this passage stand out. Volterra and d'Ancona recommend the example of mathematical physics and its successes as a template

[7]"D'ailleurs s'il apparaît trop difficile d'effectuer l'étude quantitative par voie d'expérience et d'obtenir ainsi les lois qui règlent les rapports interspécifiques dans les associations biologiques, on pourra tenter de découvrir ces mêmes lois par voie déductive et de voir ensuite si elles comportent des résultats applicables aux cas que présente l'observation ou l'expérience."(Volterra and D'Ancona, 1935, p. 8)

[8]"D'autre part, il ne faut pas trop se préoccuper si on envisage des éléments idéaux et l'on se place dans des conditions idéales qui ne sont pas tout à fait ni les éléments ni les conditions naturelles. C'est une nécessité et il suffit de rappeler les applications des mathématiques à la mécanique et à la physique qui ont amené à des résultats si importants et si utiles même pratiquement. Dans la mécanique rationnelle et dans la physique mathématique on envisage en effet les surfaces sans frottement, les fils absolument flexibles et inextensibles, les gaz parfaits, etc. L'exemple de ces sciences est un grand exemple que nous devons avoir toujours présent à l'esprit et que nous devons tàcher de suivre." (Ibid., p. 8)

for other sciences, especially biology. However, this is not an unqualified endorsement. There are reasons to adopt mathematical methods beyond the fact that they have worked well in the case of classical mechanics. The authors are acutely aware of the dangers and pitfalls of modeling.

Secondly, Volterra and d'Ancona point out that "ideal elements and conditions", what we would call idealizations, are an integral part of physics, and the success of physics suggests that the same should be tried in biology. Their emphasis on the issue of idealization should be taken seriously, as they discuss the issue not only in the methodological reflections, but over and over again, as we have seen in Volterra's 1928 account of the predator-prey model.

We believe that the reflections on the use of (mathematical) modeling is potentially due to criticisms, such as Pearson's, and that they serve as a justification of the modeling approach: if mathematical modeling is basically the only viable method, as we *cannot* separate internal and external causes in this target system, the approach gets more attractive. Volterra's and d'Ancona's position on the use of mathematical methods lies somewhere between the extremes we saw in the last section. The use of mathematics is not a means to legitimize population ecology, but grows out of the internal logic of the biological system under scrutiny; on the other hand, they fend off sweeping criticism against the use of mathematics by appealing to the authority of physics.

Of course, it is impossible to reach a verdict on the predator-prey model based on this methodological apology. However, what is the status of the model today? I turn to this question in the next section.

## 3.5 The Predator-Prey Model Today

Today, there appears to be a consensus in the literature on mathematical modeling in biology that the original predator-prey model fails to genuinely capture interactions in real biological systems. The failure is *prima facie* attributed to a mathematical property of the system of differential equations, the *stability* of the system: the mathematical system is unstable in that if there is a small disturbance of the system, this can have a big effect later on.[9]

The phase space diagram in figure 3.1 above illustrates this point. Assume we are on orbit $\Lambda$ below point $\Omega$, and the system is disturbed such that we are now on orbit $\Phi$. If we follow the two trajectories, we see that the two orbits diverge over time; the gap has markedly widened in the upper right

[9]See e.g. Murray (1993, p. 65). The predator-prey model is *neutrally stable*: if disturbed from its initial orbit, it will not return to this initial orbit, but stay on a new orbit. Neutral stability should be contrasted with *unstable* systems, who behave chaotically, and *asymptotically stable* systems, where, roughly, the disturbance gets smaller over time such that the system returns to its undisturbed state.

corner of the diagram. The property that small changes at one point can have large effects later is commonly taken to be responsible for the model's inapplicability.

At times, the formulations in the literature suggest that the problem with the predator-prey model is exclusively due to this mathematical property. However, this would not be satisfactory: If the real system in the world also showed the lack of stability we find in the equations, we would have no reason to reject the model. It is only possible to reject the model based on this mathematical property together with some empirical findings.

The following passages from a textbook on mathematical modeling in biology gives a more satisfactory explanation:

> [I]t turns out that there are serious flaws in the model. Any attempt at refinement by introducing self-limiting terms in the per capita growth rates such as in the logistic equation for single populations will lead to qualitatively different behavior of the solutions, orbits that spiral in towards the equilibrium rather than periodic orbits. The price of refinement of the model is loss of agreement with observation. (Brauer and Castillo-Chávez, 2001, pp. 129)

The problem with the Lotka-Volterra model described in this passage is that, if we remove the idealization that the prey population grows indefinitely in isolation, the model will no longer exhibit the behavior we expect, namely the periodic fluctuations, which, in turn, also affects the Third Law, as we use the periodicity of fluctuations in its derivation.

The problem described in the following passage is even more general and devastating:

> [T]he Lotka-Volterra system [...] unrealistically predicts population oscillations that have been observed in real populations [...]. The reason for describing this prediction as "unrealistic" is that the model is extremely sensitive to perturbations. A change in initial population size would produce a change to a different periodic orbit, while the addition of a perturbing term to the system of differential equations could produce the same type of change or could produce a *qualitative* change in the behavior of orbits, which might either spiral in to an equilibrium or spiral out from an equilibrium. (Ibid., p. 180)

The problems here is also due to the instability of the system, but it is wider in scope in that *any* attempt to de-isolate the system by allowing small disturbances will either lead to uncontrolled oscillations, and not the regular behavior of the unperturbed system, or to an overall qualitative change in the system. This renders the model virtually inapplicable, as there are always external factors such as seasons and migration that interfere with the system.

This, however, does not undermine the value of the model altogether. As Murray points out in his classic textbook:

> [T]his model has serious drawbacks. Nevertheless it has been of considerable value in posing highly relevant questions, is a jumping-off place for more realistic models and is the main motivation for studying it here. (Murray, 1993, p. 64)

In sum, while the original predator-prey model is ultimately flawed, because we cannot de-isolate it without changing its qualitative behavior, it is valuable as a template for more sophisticated models with more realistic growth and interaction terms. It is interesting to note that Volterra himself anticipated the possibility that there might be a problem with this particular model when he suggested that his derivation does not depend on the functional form of its interaction terms; see section 3.2.2 above. However, he does not specify which of his results depend on the structure of the model. We will scrutinize the dependence of the Third Law on the model in section 3.6.3, and we will turn to the philosophical discussion of robustness in section 3.7.2.

## 3.6 Philosophical Lessons from History

In the preceding sections, we have prepared the ground for a philosophical analysis of the predator-prey model, to which we now turn. The goal is to get a clearer picture of the role of mathematics in the construction of the model, and in the subsequent derivation of the Third Law.

We focus on four aspects of Volterra's 1928 discussion. First, we will sketch an interesting parallel between Volterra's characterization of the role of mathematics and a recent account of the applicability of mathematics, the Inferential Conception. Second, we examine the influence of mathematics on the various idealizations of the model. Third, we discuss the derivation of the Third Law, and its dependence on the specifics of the model. Fourth, we scrutinize the relation between the mathematical phase spaces and several notions of intervention.

### 3.6.1 Volterra's Account of Applicability

There is a striking similarity between Volterra's account of the role of mathematics in the application to population ecology presented in section 3.2.1 above, and an account of applicability, the Inferential Conception (IC), proposed in Bueno and Colyvan (2011), which is discussed in-depth in chapter

7. In particular, it is noteworthy that Volterra emphasizes the inferential possibilities provided by the theory of differential equations.

Volterra's stages of the application process align with the stages of the IC as follows: The first step, the abstract description of the target phenomena, corresponds to the initial assumed structure, the starting point of the modeling exercise. The second step, the translation of the informal description into differential equations, corresponds to the immersion step. The mathematical domain is the theory of ordinary differential equations, and the predator-prey equations are the target mathematical structure of the immersion mapping. The theory of differential equations is put to work in the deduction of the Third Law. Finally, after deducing the Third Law as a purely mathematical result, it is taken to explain statistical data concerning the Adriatic in the interpretation step. *Prima facie*, we are dealing with a closed cycle – the Third Law explains the fishery statistics.

It would be fruitful to carry out a deeper analysis of the application process in terms of the IC. Here is how such an analysis might proceed. The above discussion provides us with a good understanding of the application of mathematics to this particular problem. It would be interesting to examine which parts of Volterra's mathematical deductions were already available from the theory of ordinary differential equations prior to their application to population ecology. For example, does an analogue of the Third Law exist for other target systems from physics?[10] This would make it possible to gauge to what extent the model, and the subsequent deductions, were motivated by pure mathematics – say, the simplicity of equations – which, in turn, would suggest that the discovery of the Third Law was mathematics-driven, or whether it grew out of the fishery statistics, as Volterra writes.

The deductive possibilities provided by the mathematical theory, however, are only one way in which mathematics may have been the driving force behind the construction of the model. Some passages of Volterra's account suggest that the construction of the predator-prey model itself is based on mathematical considerations rather than real-world concerns. This is the topic of the next section.

### 3.6.2 The Model: Mathematics-Driven Idealizations

We discerned two different kinds of idealizing assumptions relevant for setting up the model.[11] First, Volterra restricts the phenomena that are supposed to be captured by the model; he excludes factors that are not due to the interaction of the two species or their intrinsic growth and death rates. This can be interpreted as (causal) *isolation*: only a subset of all the causes

[10] I thank Raphael Scholl for suggesting this question.

[11] The debate on idealization in models is very large by now, going at least back to Cartwright (1983). A useful overview of idealization in the context of modeling can be found in Frigg and Hartmann (2012, Sec. 1.1., 5.1.).

that influence the two populations are taken into account. While this is an idealizing assumption for most systems, the idea could be that this is harmless, as other causes can be added to the model at a later stage.

The second aspect of idealization is how to deal with those factors that *are* part of the model. Here we have identified several idealizing assumptions, such as continuous variables and continuous birth and death rates. These idealizations are of a different type in that it does not seem possible to remove them at a later stage. It is simply a fact that fish populations are discrete, that they have age structure, and so on. However, this need not render the model unrealistic. First, the description of discrete quantities with continuous variables could be a good *approximation*, or, alternatively, that the model is a good description of, say, populations with variable birth and death rates because it captures *averages* of these quantities.

However, it is questionable whether it makes sense to establish that the continuous case is an approximation of the discrete case, for the following reason. In order to show this, we would need a discrete model in order to evaluate our continuous model. Once we have a discrete and reliable model, we no longer need the continuous model. The same reasoning applies to the interpretation of the idealization as an average. Defending these idealizations based on a more realistic model is not a viable option.

We suspect that Volterra's implicit account of model building – he first discusses idealizations, and introduces the terms of the model later – does not mirror their actual role in discovery. It is more likely that this is a case of "reverse engineering": Volterra wanted to find a model in which both species evolve in a certain way if the other species is not present, and then he wanted to capture interactions between species. The predator-prey model is just the simplest possible model of this sort: both species grow, die, eat, and are eaten at a constant rate. If this is correct, then Volterra began the construction of the model by thinking about the interpretation of constant growth factors and interaction terms, and found the various idealizations after setting up the model.

This interpretation is further supported by Volterra's discussion of the interaction terms, where he comes close to admitting that what is responsible for the model's form is mathematical simplicity; a more substantive justification followed only after the fact. The fact that Volterra justified the model based on simplicity at first, and added other justifications later, creates a tension. Which justification should we take seriously? The appeal to mathematical simplicity, in the construction of models, can undermine a realistic interpretation, as it is unclear why a simple or elegant model should track truth.

Volterra appeals to the probable number of encounters as a remedy of this flaw. However, on closer inspection, the argument is not convincing. It only shows that the different constants in the interaction terms are compatible with an equal number of encounters for both terms, which is reasonable, as the situation is symmetric (if I meet you, you meet me). The difference is then explained by different conversion rates, i.e. differences in the means of offense and defense between the two species. But this does not explain why the conversion rates should be constant instead of any other functional dependence.

As a second justification, Volterra notes that the "mode of integration" he uses in the deduction does not depend on the particular form of interaction terms. This is an interesting observation, as it seems to mesh well with the ongoing philosophical debate on robustness: the idea that if we can vary features of a model, such that inferences from the model still go through, is a sort of confirmation of the model's properties.

While showing that certain results do not depend on certain specific assumptions of a model is certainly valuable, it does not empirically confirm the original model in any way, as it is the very point of a robustness result that the deduction is, to a certain degree, *independent* of said model. Therefore, a robustness result does not speak in favor of a realistic interpretation of the predator-prey model. We will further discuss the issue of robustness below.

In sum, we conjecture that, while the motivation for the predator-prey model was prompted by fishery statistics, the driving force behind the construction of the model, and especially its particular form, was mathematical simplicity, and not any real-world consideration. This speaks against a realistic interpretation of the model, as Volterra does not seem to succeed in providing good justifications for the idealizations underlying the model. It does, however, not speak against Volterra's framework.

### 3.6.3 The Derivation of the Third Law

Volterra's mathematical analysis of the predator-prey model has a noteworthy feature. He does not examine explicit solutions of the system of coupled differential equations, because it is *impossible* to find analytical solutions to the equations.[12] He only derives some qualitative features of the model, which do not depend on quantitative details; most importantly, the solutions are closed orbits of constant period.

This, however, does not mean that the results derived from this qualitative analysis are approximative in nature: Volterra's Third Law is derived independently of explicit solutions of the Lotka-Volterra equations, but it is nevertheless a precise result that holds for all orbits of the system, not only for those close to the equilibrium point; Volterra writes that the Third Law holds as long as the disturbance term $\alpha$ does not change the sign of the $\epsilon$ parameters.

Also, the Third Law depends on the specifics of the model. In particular,

[12] See e.g. Brauer and Castillo-Chávez (2001, p. 128).

the fact that the average population sizes over a period correspond to the equilibrium point, expressed in equations 3.12 - 3.13, depends on the fact that the solutions of the predator-prey equations are closed orbits of constant period. The derivation uses the fact that equation 3.9 is zero, which is only possible if both $N_1$ and $N_2$ are periodic functions of time.

These two features of the derivation, that the Third Law is not an approximative result, and that it crucially depends on the (qualitative) properties of *this* model, have been neglected in the philosophical discussion of the model, as we will see.

### 3.6.4 Phase Spaces and Their Interpretation

One of Volterra's mathematical tools in the derivation of the Third Law is the analysis of phase spaces, in particular how they are affected by changes in parameters. An interpretation of phase spaces, in the context of population dynamics, brings interesting aspects of the interaction between mathematical representation and real-world phenomena to the fore. I propose to interpret the different phase space diagrams used by Volterra in terms of different kinds of interventions in predator-prey systems. This will give us useful insights into the scope and nature of the Third Law. Note that I do not discuss the issue of a *realist* interpretation of phase spaces, but rather the question of how certain aspects of phase space suggest different notions of intervention in the system.

First, we can distinguish between a) interventions within one phase space and b) interventions that change the phase space.

For case a), consider figure 3.1 above. Usually, we think of the different orbits in phase space as corresponding to different trajectories due to different initial conditions. We can implement a change of orbit with a "surgical", one-time intervention that changes the total number of one or both of the species in a system. Assume we are on orbit X somewhere in the lower right corner. Now we instantaneously change the number of species as follows: we add a certain quantity of $N_1$ and take away a comparable quantity of $N_2$, without tampering with the reproductive abilities of the two species. The result of this intervention is that we are in the lower right corner on the new orbit $\Psi$. If we do not intervene further, the system will stay on this orbit. More importantly, we have *not* changed any of the parameters, and the average number of both species is unaffected. We are on a different orbit, but in the same phase space.

The same is not true in case b), see figure 3.2. Interventions underlying the transition from orbit $\Lambda$ to $\Lambda'$ are more complicated to realize. It is not sufficient to change the number of species at one instant of time only – this would just reset the system on a different, disjunct orbit. There are two ways in which this kind of intervention can be implemented.

We can i) change the $\epsilon$ parameters, that is, we change the natural rate of birth or death of the two species by, say, introducing a more fertile species, or shortening the life span of the animals with chemicals. This changes the parameters permanently such that we make a transition from one phase space to another: different parameters correspond to different phase spaces. This intervention can also be instantaneous (at point $P$ in figure 3.2), but it does not directly affect the number of species, just their reproductive capabilities.

Option ii) is to implement the transition from $\Lambda$ to $\Lambda'$ by *continuously* increasing or decreasing the number of one or both species, starting at $P$. This intervention has to be continuous because, presumably, adding or subtracting a certain number of animals to the population will not affect their natural reproduction rates, at least within certain limits. If the intervention is suspended, the system returns to its old orbit with its old stationary state.

Volterra thought that fishing is an instance of this kind of intervention – a certain number of animals of both species is subtracted in a continuous way. This is what motivated the whole inquiry, because a pause in the continuous intervention during WWI resulted in the qualitative changes explained by the Third Law.

However, it is not easy to implement option ii): The continuous intervention has to add or subtract a *constant* fraction of predators and prey *at all times*. We aim for a certain constant deviation $\alpha$ from the natural $\epsilon$ parameters, resulting in a trajectory in a new phase space as in figure 3.2. In order to implement this change, we have to keep the system on a new set of parameters $\epsilon_1^* = \epsilon_1 - \alpha$ and $\epsilon_2^* = \epsilon_2 + \alpha$.

If the intervention $\alpha$ is not proportional to the number of species $N_1$ and $N_2$, which vary over time, we will not stay in one phase space, but continuously wander through different phase spaces corresponding to the varying growth rates. If we stop continuously subtracting a proportion $\alpha$ of both species, the system will instantaneously go back to its old phase space with natural parameters $\epsilon_1$ and $\epsilon_2$.

This makes it impossible to bring about a change in parameters by continuous intervention in a quantitatively precise manner, because this would require that we know the functions $N_1$ and $N_2$, i.e. the solution to the predator-prey equations, and this solution we do not have. Thus, any continuous intervention, such as fishing, is necessarily an approximation of the exact change $\alpha$ in parameters described above, because we cannot know the exact number of predators and prey to be subtracted at any time. Volterra's principle is only qualitatively instantiated by fishing in the Adriatic.

In sum, there are two main lessons to be learned. First, we can interpret the different transitions between orbits in phase space diagrams as different kinds of interventions. We distinguished three kinds of interventions: a one-time change in the number of species; a change of parameters, or reproductive capabilities, and continuous intervention on the number of animals, in particular subtraction of one or both species. The second and the third kind of intervention are qualitatively on a par.

Second, the Third Law is only approximately instantiated by fishing in the Adriatic, as it is only approximately a continuous intervention. This indicates that mathematical simplicity is the driving force behind the mathematical formulation of continuous interventions. Assuming that a constant portion of each species is subtracted at every instance in time is just the simplest possible mathematical implementation of a continuous intervention in the predator-prey system. It is, however, not a very natural implementation and consequently only yields a qualitative understanding of the change of the system. A change of reproductive capabilities of the two species due to, say, some pesticide, is a much more natural real-world interpretation of a change $\alpha$ in the $\epsilon$ parameters.

This is not to downplay Volterra's achievement. Reading a change in parameters in terms of a continuous intervention, yielding a qualitative understanding of the original explanandum, is a surprising, and brilliant, conceptual reinterpretation of the change of parameters.

## 3.7 The Predator-Prey Model in the Philosophical Discussion

In this section, we look at some recent discussions of the predator-prey model in the philosophical literature. Our discussion of Volterra's historical account of the model will permit us to clarify some (mis-)conceptions in these debates, and to sharpen our systematic understanding of the case.

### 3.7.1 Colyvan on Population Ecology

Mark Colyvan (2013) discusses the Lotka-Volterra model in the context of the philosophical debate on the applicability of mathematics in the sciences.[13] He thinks that, to this day, the use of mathematical methods in the *special sciences* has been challenged, if not outright rejected. His goal is to establish that mathematical models in the special sciences are not only helpful for prediction, but that they can play a genuinely explanatory role. This claim is supported by various case studies from biology; one of them is the predator-prey model.

I agree with Colyvan's main thesis that mathematical models in general, and models in population ecology in particular, can be explanatory, and that there are good reasons to accept the use of mathematical models in the special sciences – despite the fact that the original predator-prey model probably fails to be explanatory. Colyvan's main thesis can be supported historically – explanatory concerns played a key role in the genesis of mathematical population ecology.

[13] This section draws on Räz (2013a).

As we saw in the beginning of section 3.2, the predator-prey model was not constructed merely to reproduce a certain phenomenon, but to explain a high-level feature of a system of fish populations. The project of mathematical modeling in population ecology was explanatory from the start. What is more, the mathematical model only gives us an approximate, qualitative understanding of the population interactions, as the coupled, nonlinear differential equations generally cannot be solved analytically. The material motivation for the model, as well as its mathematical features, lead to a qualitative understanding of the system under scrutiny – precise quantitative predictions are neither the motivation behind the model, nor are they possible.

On a more critical note, I think that Colyvan does not sufficiently emphasize the issue of idealization, one of the main problems of (mathematical) modeling. In his response to the objection that mathematical models fail to give an adequate account of ecological systems, he distinguishes between the claim that mathematical models fail to represent at all, and the claim that mathematical models are overly simple, or misrepresent; see Colyvan (2013, p. 4). He refutes the first claim, by pointing out that the parameters of the predator-prey model can be interpreted as summing up biological information, such as birth and death rates.

In his response to the second claim, he grants that the original predator-prey model is overly simple, but points out that its role in modern population ecology is pedagogical, and that it can serve as a template for more sophisticated models. He thinks that even these models leave out biological details, which, however, does not invalidate them, as leaving out details is part and parcel of the practice of modeling. In sum, Colyvan considers the first claim to be true, but somewhat beside the point.

I agree with Colyvan's characterization of the model's role in modern population ecology (see the quote at the end of section 3.5), and that the predator-prey model does sum up and represent biological information. The model's parameters have a clear biological correlate. However, I find the claim that (some) mathematical models, such as the predator-prey model, are overly simple and misrepresent to be much more interesting, and troubling. The real challenge of mathematical population ecology is to tell a story about how some models that lie about their target system can nevertheless be explanatory, while others are worthless.

The claim that the issue of idealization is of utmost importance is supported by a close reading of Volterra's and d'Ancona's original publications. The use of idealizations is acknowledged and defended throughout Volterra (1928) and Volterra and D'Ancona (1935), as we saw in sections 3.2 and 3.4. Volterra and d'Ancona were aware of the fact that they made ample use of "ideal elements and conditions". They defend this practice based on general considerations, such as the appeal to the use of idealizations in physics, but also by attempting to justify particular idealizations, such as the use of continuous variables and constant growth coefficients. Idealization is one of the problems Volterra and D'Ancona took seriously.

In sum, Colyvan's claim that mathematical models in the special sciences have an important, and even explanatory role, is supported by historical facts; population ecology relied on mathematical models for explanatory purposes from the beginning. I urged that one of the most important problems with mathematical models is idealization; this should be emphasized more. However, this does not invalidate Colyvan's general philosophical claims.

### 3.7.2 Weisberg and Reisman on the Volterra Principle

Weisberg and Reisman (2008) discuss the predator-prey model in the context of so-called *robustness analysis* in population ecology. The idea behind robustness analysis is the following. Mathematical models in biology are often highly idealized, and it can be unclear whether properties of a model are generic or an artifact of the mathematics used in the formulation of the model. To prevent the danger of mathematical artifacts, modelers employ the technique of robustness analysis, which involves the construction and analysis of multiple models. The hope is that if multiple, independent models are able to, say, reproduce a phenomenon, this is taken to indicate that this phenomenon is a genuine feature of these models, and not due to idealizations used in any one of them. The slogan of robustness analysis goes back to Richard Levins, who wrote that "[O]ur truth is at the intersection of independent lies" (quoted after Weisberg and Reisman (2008, p. 107)).

Weisberg and Reisman detail three ways in which a mathematical model can be robust. First, *parameter robustness* investigates whether features of a model depend on particular settings of the parameters of a models (think of settings of the $\epsilon$ and $\gamma$ parameters in equations 3.2 and 3.1). Second, *structural robustness* is a more severe test in that different models of a certain general kind are compared: the structure of the models comes into focus. For example, models with different interaction terms could be compared, while the modeler would rely on the framework of (coupled) ordinary differential equations. Third, *representational robustness* is a further generalization in that different models from different mathematical frameworks, such as models with continuous and discrete time steps, are compared.

Weisberg and Reisman put these categories to work in population ecology; they compare several mathematical models, among them the original predator-prey model. One of their main results is that, what they call the *Volterra Principle*, is a robust result in that it has all three kinds of robustness for a range of predator-prey models. The Volterra Principle states that "a general biocide, any substance which has a harmful effect on both predators and prey, will *increase* the relative abundance of the prey population" (Ibid., p. 113, emphasis in original).

I will argue in the following that several of the details in Weisberg's and Reisman's analysis of the original predator-prey model are historically and systematically inaccurate, and that this invalidates their main thesis, viz. that the Volterra Principle is a robust property of predator-prey models. Volterra's Third Law is arguably the most important result of Volterra (1928), and this is what Weisberg and Reisman refer to as the Volterra Principle. However, the Third Law is *not* a structurally robust property of predator-prey models. Here is why.

Weisberg and Reisman first discuss parameter robustness of some properties of the original predator-prey model. They note that the model shows undampened oscillations and is *neutrally stable*: if the system is undisturbed, it will oscillate on a closed trajectory, and if it is disturbed, it will make a transition to a different trajectory, and stay on this new trajectory until it is further disturbed. More importantly, the model has the *Volterra Property*, the key component of the Volterra Principle: if a general biocide is applied, i.e. if both predators and prey are destroyed, the proportion of prey relative to predators is increased.

I have two clarificatory remarks on this. First, the biocide has to be of the right kind: it is necessary to continuously destroy both species, or to affect their reproductive capabilities, as we saw in section 3.6.4 above. Secondly, parameter robustness has to be formulated carefully in the case of the Volterra Property, as the property itself relies on a variation of parameters. Volterra himself analyzes parameter robustness when discussing the scope of the Third Law in terms of the range of $\alpha$, see section 3.2.4.

Weisberg and Reisman then turn to the analysis of the structural robustness of these properties. To this end, they introduce a modified predator-prey model with density dependence: The linear term of equation 3.1 is modified such that, if left alone, the prey population would not grow indefinitely, but only to a certain maximum, the so-called carrying capacity. This model has no, or only dampened, oscillations and is not neutrally stable, as the populations converge to one of three equilibrium points depending on initial conditions. These properties are not structurally robust. Weisberg and Reisman also claim that the model with density dependence still has the Volterra Property.

I think that this last claim is historically incorrect and in need of systematic clarification. We saw in section 3.2.4 above that in the derivation of the Third Law, Volterra used the fact that the trajectories of the original predator-prey model are closed cycles of constant period; specifically, we need this property in order for equation 3.9 to be equal to zero. However, this means that if a model has only dampened, or no, oscillations, as the model with density dependence, then we also cannot derive the Third Law in Volterra's manner. Put differently, if the Volterra Principle is the Third Law, then it is impossible that undampened oscillations and neutral stability are not structurally robust, but the Third Law is, as the latter depends on the former.

What has gone wrong in Weisberg's and Reisman's analysis of the Volterra Principle a.k.a. the Third Law? There are crucial differences between their account of the Volterra Principle and Volterra's account of the Third Law. Weisberg's and Reisman's analysis of the Volterra Principle begins on p. 113. They write that the equilibrium point $\Omega = (\epsilon_2/\gamma_2, \epsilon_1/\gamma_1)$, see figure 3.1, can be found by setting the equations equal to zero. The ratio of the two components then yields the relative abundances of the two species. Weisberg and Reisman then note that the equilibrium points "correspond to the average abundance of the predator and prey species over indefinitely long time periods" (Ibid., p. 113).

Here Volterra's deduction of the Third Law becomes relevant. Volterra gave a precise mathematical meaning to average abundances in equation 3.6. This does *not* correspond to an average abundance over "indefinitely long time periods", but to exactly one oscillation. More importantly, the argument as to why the equilibrium point corresponds to the average abundance, i.e. essentially the whole derivation of the Third Law, is missing in Weisberg's and Reisman's account. It is probably precisely because they left out this argument that they did not recognize the dependence of the Third Law on closed cycles of constant period.[14]

We can trace the origin of this less than satisfactory treatment of the original predator-prey model to the literature on population ecology used by Weisberg and Reisman. They cite Roughgarden (1997) and May (2001) in their discussion of the Volterra Principle. Here is Roughgarden's account of the relation between equilibrium point and average abundance:

> [W]e can use the formulas for the equilibrium abundances as approximate predictors of the *average* abundances through time. If the trajectories were circles, the equilibrium point would be exactly the average through time, and because the trajectories are not quite circular, the equilibrium point is only approximately the average through time. But let's look at these formulas anyway. (Ibid., p. 270, emphasis in original)

This is not quite right, as we have seen above. The equilibrium is not only an "approximate predictor" of average abundances, but the exact average abundance. Also, the result is independent of the form of the trajectories in phase space: It hold for *all* kinds of (closed) trajectories. This is clear from the mathematical derivation, which does not put any restriction on the shape of orbits.

[14]Maybe Weisberg's and Reisman's ideas could be sanitized by a) spelling out what they mean by "average abundance", say, a limit of the time average, and b) showing that Volterra's notion, as well as that of other models, falls under this notion of average abundance. I thank Claus Beisbart for suggesting this possibility.

May (2001) has a very brief discussion of the original predator-prey model which does not draw on Volterra's original work.[15] May relies on an eigenvalue analysis of the model, which, presupposes a linearization, or approximation, of the original model. This is perfectly fine, but it could undermine the generality of certain of Volterra's results; for example, the approximation may not be good enough for certain parameter values. If this is what Weisberg and Reisman had in mind, they should have discussed the ramifications of using an approximation.

In sum, Weisberg's and Reisman's account of the Volterra Principle a.k.a. the Third Law is certainly historically inaccurate, and systematically wanting, in that Volterra's Third Law presupposes closed cycles of constant period in order to be valid, and it is a general, not an a approximative result. Systematically, Weisberg and Reisman owe us an explanation as to why the equilibrium in the original model corresponds to an average of the population sizes, and if this account depends on an approximation, they should discuss how this affects the (parameter) robustness of the properties under discussion. As I pointed out above, the discussions of the relevant aspects of the model in the ecology literature is often either very brief, or inadequate.

Let me conclude with some remarks on Weisberg's and Reisman's ideas and the prospects of robustness analysis in general. First, I think that the idea of identifying robust properties, such as the Volterra Property, is an interesting idea. However, much hinges on the precise mathematical formulation of the ideas involved. We saw that if vague ideas, such as that of average abundance, are mathematically disambiguated, it can happen that the robustness of a property breaks down. Once more, the devil is in the detail.

Second, reading the introductory paragraphs of Weisberg's and Reisman's paper, one can get the impression that the idea behind robustness analysis is a novel approach first proposed by Levins, and particularly useful in (population) biology, e.g. when the authors write:

> Biologists often value results that are general – for example, a theoretical treatment of a system that remains true under many possible states of the system, or a result that applies to a wide range of different systems. Recognizing that any body of theory will depend on some set of assumptions, biologists possessing a general result will often want to know whether it will continue to apply under differing assumptions about the system. (Weisberg and Reisman, 2008, p. 107)

It seems to me that everything Weisberg and Reisman say about biologists's predilections is equally true for any scientist, or mathematician,

[15] I was unable to find p. 439 of May's book indicated in Weisberg and Reisman (2008, p. 114), as the book only has 265 pages *in toto*.

working with abstract mathematical models. Whenever we work with such a model, it is crucial to understand how the results derived from the model depend on particular assumptions, and we will always strive to generalize the model and check whether results are still valid under weaker assumptions. This is a very common scientific practice, and I am not convinced that it is necessary to introduce "robustness analysis" as a novelty for biological or ecological contexts. I have not yet been able to discern an argument showing the particular value of robustness analysis in biology as opposed to, say, physics. A philosophical discussion of these issues in general would certainly be valuable.

### 3.7.3 Pincock 2012

In a recent monograph on the role of mathematics in scientific representation, Christopher Pincock (2012) discusses the predator-prey model as an example of an *acausal* representation. He gives two reasons for classifying the representation as acausal: The first is that the model misrepresents actual causal processes, by introducing various idealizations, such as constant growth, death and interaction parameters, and by neglecting influences from the environment. The second, more important reason for classifying the representation as acausal is that "the genuine causal actors in these systems are the individual animals and not anything like the number of predators or prey in the system" (Ibid., p. 59). Pincock thinks that a representation of a biological system on this level will involve "a host of inaccuracies".

Pincock maintains that the model is useful by providing an acausal representation via Volterra's Third Law. This claim is based on the alleged robustness of the Third Law as proposed in Weisberg and Reisman (2008): the predator-prey model may be an inaccurate representation for individual parameter values, but if we take a whole set of models and ranges of parameter values into account, the Third Law holds true in this set of models and range of parameters, and we have reason to believe that the result is independent of individual parameter values.

I have already discussed the prospects of robustness analysis in section 3.7.2 above. Everything hinges on the precise formulation of the concepts involved; if done properly, it turns out that the Volterra Principle, a.k.a. the Third Law, is not nearly as universal as one might have thought.

Regarding the question as to whether the predator-prey model is a causal representation or not: I think we should not require a model to be accurate, or to mirror a target system faithfully, in order for it to be a model that is supposed to mirror the causal structure of a system. Concerns of idealizations and causality can (but need not be) orthogonal. For example, if a model merely isolates core factors out of a totality of relevant causes of a phenomenon, we would certainly classify it as causal, despite its idealizing nature.

Pincock's second reason for thinking that the predator-prey model is an acausal representation, namely that the system is described at the level of populations, while the underlying causes are located at the level of individual animals, is also not convincing. First, I see no principled reason why the level of individual animals is a privileged level of description. A description of the system on the level of, say, fundamental particles could be required by the same kind of argument; however, such an explanation would be less explanatory than a higher level account. The same is probably true of a description of the system on the level of individual animals: Volterra's (and population ecology's) goal is to describe and explain population level phenomena, and taking individual fish into account just introduces irrelevant details. Secondly, as Mark Colyvan (2013) points out, parts of the predator-prey model do correspond to a causal description of the system: It seems entirely reasonable to view the birth and death rates as major *causes* of population sizes. These factors are more or less directly represented by the parameters of the model, albeit in an idealized manner.

## 3.8 Conclusions and Outlook

In this final section, I offer conclusions and note questions that could be pursued further.

First, we saw in sections 3.2 and 3.4 that in Volterra's (and d'Ancona's) account, the material motivation for proposing the predator-prey model was explanatory, namely to account for the shift in proportion of predators and prey during the first world war by invoking the Third Law. Furthermore, the discussion in Volterra and D'Ancona (1935) shows that they view modeling as an alternative account when other methods, such as direct causal inference, fail.

It is an open question as to whether the fishery statistics really were the starting point of Volterra's investigations. An alternative explanation is that the motivation for the model came directly from mathematics; more specifically, from a physical analogue of the Third Law. To substantiate this alternative motivation, we would have to search the mathematical literature on ordinary differential equations and their application to physical problems available to Volterra.

Second, I argued in section 3.6.1 that Volterra's mathematical methodology is strikingly similar to the Inferential Conception, especially in that Volterra emphasizes the deductive possibilities offered by the theory of ordinary differential equations. To fully appreciate how the application process works in this case, we need a better understanding of the relevant mathematical theories.

Third, the issue of idealization in modeling in general, and in the predatory-prey model in particular, is paramount and cannot be overemphasized. We saw in the discussion of Colyvan's contribution in section 3.7.1 that this is not always appreciated in the philosophical discussion. I discerned different kinds of idealizations relevant to the model. While the isolation from external, environmental factors seems to be driven by real-world concerns, other idealizations apparently have their origin in mathematical simplicity. We would expect the latter kind of idealization to be in need of justification; however, this is not accomplished by Volterra.

Historically, idealizations were important in the discussion of the model from the start; systematically, the ultimate reason for the rejection of the model is the impossibility of de-idealizing the model without changing its qualitative features. However, the model is still important as a template for more realistic and complicated models. An open question remains as to the evolution of the idealizations and their assessment by population ecologists and mathematicians. For example, what was the path to the discovery of the structural instability of the system?

Fourth, I discussed several instances where the mathematical details of the model can inform and improve philosophical debates. One example is the derivation of the Third Law, which, in its original form, is a general, not an approximate result, and one which depends on other properties of the predator-prey model, in particular that the phase space consists of closed orbits with constant period. This fact has been neglected in philosophical discussions, which has led to misconceptions about the robustness of the Third Law. A second example is the connection between phase spaces and intervention: I discerned several notions of intervention that are suggested, at least in part, by the mathematics. They are relevant to the application of the model, as only the right kind of intervention will lead to a shift of population averages as predicted by the Third Law. It could be interesting to explore whether these distinctions can be used in philosophical debates, say, on interventionism. I argued in section 3.7.2 that discussions of issues like robustness should be conducted at the general level of mathematical modeling; there is nothing particularly biological about robustness.

Fifth, we touched on the issue of causality. I argued in section 3.7.3 that the model has a direct causal interpretation in that the parameters can be interpreted causally: growth rates seems to be a major cause of the size of populations. We also argued that modeling can be interpreted as an alternative to a more direct approach involving causal inferences.

# Chapter 4

# The Bee's Honeycomb

## 4.1 Introduction

In this chapter, I discuss scientific and mathematical attempts to explain the structure of the bee's honeycomb. The chapter has two parts. In the first part, I discuss a candidate explanation, based on the mathematical Honeycomb Conjecture, a standard example in the philosophical debate on mathematical explanations of physical phenomena. I argue that this explanation is not scientifically adequate. I will cast doubt on the idea that the Honeycomb Conjecture is part of an explanation of the structure of the bee's honeycomb – the purported explanation is flawed on mathematical grounds.[1]

In the second part, I discuss other mathematical, physical and biological studies that could contribute to an explanation of the bee's honeycomb. The upshot is that most of the relevant mathematics is not yet sufficiently understood, and that there is an ongoing debate about the biological details of the construction of the bee's honeycomb. There are, however, relevant and promising results from the physics of foams that, depending on the outcome of the biological debate, could provide an explanation.

### 4.1.1 Overview

Here is a section-by-section overview of the chapter.

In section 4.2, I give a brief account of the explanation of the structure of the bee's honeycomb based on the Honeycomb Conjecture (HC) as presented in Lyon and Colyvan (2008).

In section 4.3, I establish the importance of the example for a recent philosophical argument by Alan Baker (2012).

In section 4.4, I argue that the explanation proposed by Lyon and Colyvan (2008) is inadequate for two reasons. I first show that the explanation is deficient because the HC solves a two-dimensional problem, whereas an

[1] This part of the chapter is a version of my paper Räz (2013b).

actual honeycomb has a three-dimensional structure that cannot be adequately captured in two dimensions. This establishes that the HC provides only a fraction of the mathematics relevant to the bee's honeycomb. I then cast doubt on the idea that we should accept the HC even as a partial explanation of the actual, three-dimensional honeycomb. The problem is that once we consider the honeycomb in three dimensions, the adequacy of a separate explanation of the two-dimensional hexagonal substructure becomes dubious.

In section 4.5, I examine a different mathematical explanation of the bee's honeycomb proposed by Laszlo Fejes Tóth (1964), which has hitherto been neglected in the philosophical discussion. I argue that the mathematical result of this account is not applicable to the bee's honeycomb because one of the idealizations it introduces is too strong.

In section 4.6, I first review some recent biological investigations of the bee's honeycomb. I then call attention to an alternative explanation of the bee's honeycomb based on these results. This alternative explanation does not rely on the HC and thus further undermines the latter's relevance.

In section 4.7, I first introduce a general framework that helps us systematize all candidate explanations discussed so far. This framework classifies the bee's honeycomb as a kind of foam. I then give a short account of an experiment that can be interpreted as a physical realization of the bee's honeycomb; the analogy between foams and honeycombs is partially corroborated by the biological results discussed in section 4.6.

I conclude in section 4.8, with an emphasis on the philosophical ramifications of the scientific and mathematical results.

## 4.2 Lyon's and Colyvan's Explanation

The explanation of the geometric structure of the bee's honeycomb, based on the Honeycomb Conjecture (HC), was first proposed by Aidan Lyon and Mark Colyvan in their 2008 paper "The Explanatory Power of Phase Space".[2] The explanandum is that the bee's honeycomb has a hexagonal shape, as opposed to some other geometric shape. The explanans has two parts, one biological, the other mathematical. The biological part is that it is evolutionary advantageous to minimize the amount of wax used in the construction of honeycombs; Lyon and Colyvan trace this part of the explanation back to Darwin. The mathematical part of the explanation is provided by the Honeycomb Conjecture and its recent proof by Thomas Hales (2001). The Honeycomb Conjecture states that "*a hexagonal grid represents the best way*

[2] This is not to say that Lyon and Colyvan are the first to suggest a connection between the bee's honeycomb and some mathematical conjecture. As early as 36 B.C., Marcus Terentius Varro claimed that the hexagon "encloses the greatest amount of space", which explains the structure of the bee's honeycomb; see Hales (2000, p. 448). However, Lyon and Colyvan introduced the explanation into the philosophical discussion.

*to divide a surface into regions of equal area with the least total perimeter*"(Lyon and Colyvan, 2008, pp. 228). Figure 4.1 shows a part of the hexagonal grid. Lyon and Colyvan claim that the combination of these facts explains the structure of the bee's honeycomb.

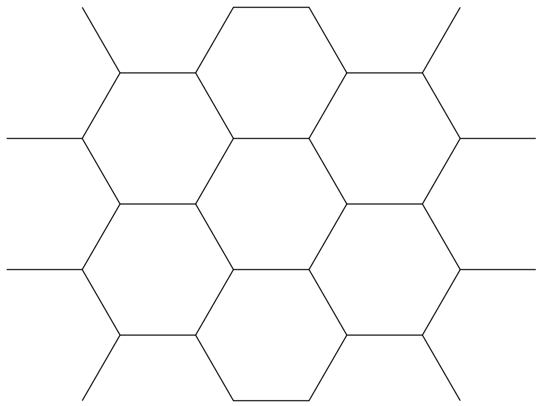


Figure 4.1: Hexagonal Tiling

This explanation of a physical phenomenon based on a mathematical theorem has subsequently been adopted as a standard example in the philosophical discussion of mathematical explanations; see e.g. Baker (2009), Baker and Colyvan (2011), Saatsi (2011), Lyon (2012), Baker (2012), Tallant (2013) – and there has been considerable disagreement about its philosophical analysis and significance.

So far, it has never been disputed that the explanation given by Lyon and Colyvan is acceptable on mathematical or scientific grounds. However, this is what I will to do in section 4.4, after first illustrating the importance of the case for the philosophical discussion.

## 4.3 Baker: A Philosophical Motivation

One might think that the honeycomb case is but one of many examples proposed and discussed by philosophers, and that therefore, while it is regrettable if it turns out not to be an actual explanation, this will not really affect philosophical arguments. However, this is not so. A recent paper by Alan Baker (2012) relies to a large extend on the scientific adequacy of the honeycomb case.

In his paper, Baker attacks the so-called *Transmission View* of Mathematical Explanation in Science (MES), which he attributes to Mark Steiner (1978b). According to this view, MES works via a transmission of an intra-mathematical explanation to some physical explanandum. The MES with explanans $M$, typically a proof, used in the explanation of a physical explanandum $P^*$, written $M \to P^*$ is, first and foremost, an explanation of an intra-mathematical explanandum $M^*$, written $M \to M^*$, and the explanation of $M^*$ is transmitted to $P^*$ via a bridge principle, written $M^* \leftrightarrow P^*$. If we remove the bridge principle from the complete MES, $M \to M^* \leftrightarrow P^*$, we are left with an intra-mathematical explanation, $M \to M^*$.

According to Baker, there are two separate problems with this view. The first is a counterexample, the honeycomb case. The second is an argument for the thesis that the proof of a mathematical theorem is not necessarily part of a scientific explanation, even if the theorem is used in that explanation.[3]

Baker notes that two conditions have to hold in order for the honeycomb case to be a counterexample. First, it has to be a genuine MES, and second, the proof must not explain the theorem. Baker thinks that the honeycomb case is clearly a genuine MES. He writes that

> [T]here is not much to be said [on this condition], other than that biologists do generally take this to be the best explanation of why honeybees build their cells in the shape of hexagons, and that it clearly makes nontrivial use of mathematics (Baker, 2012, p. 250).

Baker repeats the claim that biologists take this explanation seriously later in his paper, but he does not substantiate it with references.

I will argue below that the first condition does not hold. If this is so, then Baker's main counterexample is flawed. Thus, the honeycomb case is worth our attention.

## 4.4 Why the Explanation Fails

### 4.4.1 The Explanation is Incomplete

Lyon and Colyvan claim that Hales's theorem can help to explain what they call the hexagonal structure of the honeycomb. This presupposes that the structure of the honeycomb is in fact hexagonal – but this is incorrect. It is only the form of the *openings* of honeycombs, or their prismatic base, that show a hexagonal pattern, not the entire honeycomb. Actual honeycombs show hexagonal openings on the surface, but their actual geometric structure is more complicated than this: honeycombs consist of two layers of congruent cells, each one with a hexagonal opening and a non-flat bottom; see figure 4.2 for an approximate geometrical representation of a cell.

Mathematically speaking, the problem with Lyon's and Colyvan's proposal is that their explanation applies to a two-dimensional structure, whereas

[3] Baker does not want to rest his entire argument on just one counterexample, and notes the application of the four-color theorem as a second counterexample. I will not assess the strength of this example. However, the honeycomb case is his main case study and appears to carry most of the argumentative weight. I will discuss Baker's second problem in chapter 2.

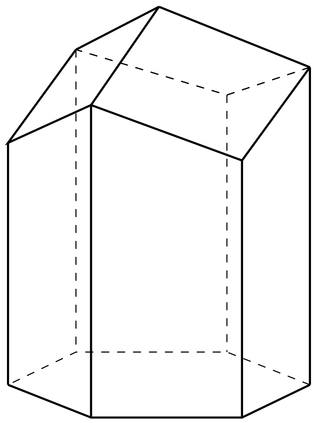


Figure 4.2: The three-dimensional honeycomb

the actual honeycomb is three-dimensional. The HC says: "any partition of the plane into regions of equal area has perimeter at least that of the regular hexagonal honeycomb tiling" (Hales, 2001, abstract). The bee's honeycomb, however, is a three-dimensional structure that does not reduce to the two-dimensional case. What we should be looking for is an optimal three-dimensional structure that can be applied to the actual honeycomb instead of the two-dimensional HC. Put simply, the structure should minimize area relative to the volume of cells instead of perimeter relative to area.

To treat the honeycomb as a two-dimensional optimization problem is not *a priori* unreasonable, but on closer inspection, it proves to be problematic. For example, if the honeycomb consisted of one thin layer of hexagonal cells only, then a two-dimensional description would probably capture the relevant aspects of the structure.[4] However, it is simply a fact that the honeycomb has a *non-trivial* three-dimensional structure. The critical point is that the actual structure comes in two layers such that the cells are open on one end only. That is why we cannot possibly account for the shape of the rhombic caps in two dimensions: the caps do not fit into the two-dimensional representation and would have to be omitted – the third dimension is necessary to represent this aspect of the structure. The three rhombi can be seen very nicely in the geometrical representation; see the top of figure 4.2.

The problem I just raised has gone unnoticed in the philosophical discussion of the honeycomb case, but it is well-known in the mathematical literature. For example, Erica Klarreich discusses Hales's proof of the HC and writes:

> Hales's work confirms that the hexagonal arrangement is the

[4]This has been suggested by Erica Klarreich; see the quote below. The qualification "probably" is necessary because the claim that this structure is optimal would have to be proven, despite its plausibility.

> one that uses the smallest amount of beeswax to create a single thin layer of cells, open on each end. In an actual honeycomb, the cells in each layer are capped by three rhombic faces, forming a rhombic dodecahedron. (Klarreich, 2000, p. 157)

Klarreich at least implicitly acknowledges that the HC does not directly apply to an actual honeycomb. The three-dimensional structure to which the two-dimensional HC applies is a prismatic extension of the hexagonal grid, while what Klarreich calls the actual honeycomb is depicted in figure 4.2.

Some formulations in the mathematical literature are even more succinct. Frank Morgan discusses the HC in his introduction to geometric measure theory. Immediately after stating and proving the HC, he adds the following observations under the title "The Bees' Honeycomb":

> The bees actually have a more complicated, three-dimensional problem involving how the ends of the hexagonal cells are shaped to interlock with the ends of the cells on the other side. L. Fejes Tóth [...] showed that the bees' three-dimensional structure can be improved slightly, at least for the mathematical model with infinitely thin walls. (Morgan, 1988, pp. 166)

Morgan states that the hexagonal grid of the HC is not the relevant structure for the actual honeycomb, and he even mentions an alternative approach.

It could be thought that it was Thomas Hales who suggested this application of the HC – but this is not so. In his paper proving the HC, Hales discusses the historical link between the conjecture and the bee's honeycomb, but he does not advocate an application of the theorem along the lines of Lyon's and Colyvan's proposal.

There is a gap between the philosophical and the mathematical discussion. Mathematicians think of the bee's honeycomb as three-dimensional and do not attempt to explain it via HC. Philosophers, on the other hand, have unfortunately neglected the three-dimensionality of the structure to date.

Even if we disregard the shape of the real honeycomb, there is a systematic problem with the explanation based on the HC: It applies to two-dimensional surfaces and therefore can only take the shape of the openings of the cells into account. This, however, is not sufficient. It is not clear that a structure with optimally shaped openings minimizes the amount of wax. A structure can have cells with optimal openings, but some non-optimal shape otherwise. We cannot infer the optimality of cells from the shape of the openings. To make sure that a structure is optimal, we have to take the whole three-dimensional structure into account.

### 4.4.2 The Honeycomb Conjecture Is (Probably) Irrelevant

We saw in the last subsection that Lyon's and Colyvan's explanation is incomplete: it cannot capture all that is mathematically relevant about the actual honeycomb. This, however, still leaves open the possibility that the HC is of *some* relevance to the actual honeycomb. After all, the openings of the cells are hexagonal – see the bottom of figure 4.2 – so it is possible that we can apply the HC to explain the optimal shape of the openings. This would constitute a partial explanation in that the HC explains a part of the structure. In this subsection, I will argue that the HC is probably not even a partial explanation of the shape of the actual honeycomb.

The argument is not directed against the use of mathematical optimization in explanations of physical structures; some form of mathematical optimization may be relevant to the bee's honeycomb. Before I proceed, therefore, it may be helpful to clarify the role of mathematics in this kind of explanation. What is necessary for a successful explanation involving mathematical optimization?

I argued above that the relevant optimization problem is three-dimensional in the present case. What is minimized is the amount of wax relative to cells of unit volume. Then, the optimization problem has to satisfy certain boundary conditions; one of them is that each cell needs an opening of reasonable size. A possible mathematical formulation of the problem is as a bounded form of the Kelvin problem: the optimal tiling of space with cells of equal volume, with the restriction that the cells lie between two parallel planes such that each cell has an opening in one of the planes. Laszlo Fejes Tóth (1964) analyzed this kind of problem; we will turn to his proposal in section 4.5. Additionally, the optimization will probably have to take the thickness of the walls into account. Of course, purely mathematical considerations will not do. For example, we have to find out if and how an "optimization process" is implemented in the world: do the bees construct the beehive from beginning to end, or is some other process involved? These issues are still debated in the biological literature, as we will see in section 4.6 below. Finally, there are some structural constraints due to stability.

The question whether any form of mathematical optimization is relevant to the bee's honeycomb is an open scientific question. For the sake of the argument, I assume that some three-dimensional optimization problem is in fact relevant, and I answer the question as to whether, under this assumption, the two-dimensional HC is relevant to the explanation as well.

We can distinguish two cases. The first possibility is that the solution to the right mathematical optimization problem does not have cells with hexagonal openings at all. In this case, the bee's honeycomb would simply not be an optimal solution, and the HC would be inapplicable, as the hexagonal tiling is not part of the structure. This is a real possibility: Three-dimensional geometric optimization problems are notoriously hard, and optimal solutions to three-dimensional problems relevant to the bee's honeycomb are not known. To give an example, we do not know the optimal solutions to the aforementioned mathematical honeycomb structures proposed in Fejes Tóth (1964).

The second possibility is that the hexagonal grid is part of the three-dimensional structure that constitutes the solution to a three-dimensional optimization problem. Even if this is the case, it is still probable that the hexagonal grid is part of the real, three-dimensional honeycomb because this whole structure is optimal in three-dimensions, and not because the grid is the optimal solution to a two-dimensional problem. This is, once more, a qualified statement, as the relevant mathematical results for three-dimensional optimization problems are not known. The hexagonal grid is nothing but a geometrical structure, and the fact that it is part of a more complex structure may be unrelated to the fact that the hexagonal grid features in the HC.

What would have to be established to show the relevance of the HC to a three-dimensional structure? The HC is one possible explanation of the shape of openings of the three-dimensional honeycomb structure. However, the structure is three-dimensional, and it is probably the solution to a different, three-dimensional optimization problem that will explain the shape of the entire structure. So one way to establish the relevance of the HC would be to show that the reason why the three-dimensional structure has its shape subsumes the reason why a part of the structure has its shape. Or, to put it differently: the proof of the optimality of the three-dimensional structure would somehow have to imply the proof in the two-dimensional case, i.e. the proof of the HC. It is unclear whether such a relation between results is plausible or can be established, as we simply do not know the relevant optimality results. However, as long as we do not know whether the HC is relevant here, we should suspend our judgement about this case.

It could be objected that the hexagonal shape of the openings does not have to be a consequence of the optimality of the three-dimensional structure, because it is a biological requirement that the cells of the honeycomb have (two-dimensional) openings in the shape of the hexagonal grid. It would then be reasonable to postulate this structure as a kind of boundary condition for the optimality of the entire, three-dimensional cells.

If we, however, simply postulate the hexagonal structure as a boundary condition, then the HC loses its explanatory power. In this case, we do not use the HC to explain the structure of the bee's honeycomb, but we use it in the deduction of a different result, the optimality of a three-dimensional structure. If, on the other hand, we could prove that a) the hexagonal openings are part of the optimal three-dimensional cell structure, and that b) the proof of the optimality of the three-dimensional structure really subsumes the two-dimensional, hexagonal case, the relevance of the HC would not have to be postulated, but it would follow from a stronger result. In this case, the

HC would indeed be explanatory.

I have not ruled out the possibility that the HC is explanatorily relevant to the actual honeycomb. However, the arguments in this section show that the relevance of the HC depends on optimality results in three dimensions, and as we do not yet have a clue what the optimal solution in three dimensions might be, we should abstain from such speculations at this point of mathematical progress.

## 4.5 Fejes Tóth: A Mathematical Proposal

I will now turn to the second goal of the chapter, namely finding out what the best available scientific explanation of the structure of the bee's honeycomb is. I will examine scientific and mathematical results that shed light on this question.

The most important paper that examines the honeycomb from a mathematical point of view is well known in the mathematical literature[5], but it has not been taken into account by philosophers. The paper is entitled "What the Bees Know and What They do not Know" and was written by Laszlo Fejes Tóth (1964).

Fejes Tóth's goal is to find out whether the bee's honeycomb is an optimal solution to some relevant geometrical optimization problems. His most important result is that the bee's solution is not optimal. While this result is interesting within pure mathematics, it does not warrant conclusions about actual honeycombs, because it essentially depends on the problematic assumption that the walls of the honeycomb are infinitely thin. Fejes Tóth readily acknowledges this problem; I will explain his result and its limits in some detail.

### 4.5.1 Fejes Tóth's Proposal

Here is a short account of Fejes Tóth's results. He starts with a description of what he takes to be the actual structure of the bee's honeycomb. It is a structure entirely made up of one type of cell. These are hexagonal prisms with one open end and a bottom formed by three rhombi; see figure 4.2 (dubbed "Actual Honeycomb" in the following). These cells are arranged in two layers, oriented in opposite directions with the hexagonal openings pointing outwards, showing the characteristic seamless hexagonal tiling, and the bottom side of each layer meeting the other layer – the landscapes made up of three-rhombi-hills neatly fit together.

Fejes Tóth's goal is to compare this and other possible designs with respect to their surface area: is the bee's structure really the structure that

[5]See the quotes by Klarreich and Morgan above, as well as Hales's paper on the HC, for remarks on Fejes Tóth in the mathematical literature.

minimizes surface area, or do other, superior designs exist? The biological motivation for his approach is that there is a direct correspondence between the surface area of a given structure and the amount of wax necessary to build a honeycomb: less surface area means that less wax is necessary to build a honeycomb. Of course, it could be asked whether surface is a good measure for the amount of wax necessary to build a honeycomb. We will return to this point in the discussion of Fejes Tóth's proposal.

Before one can even start to compare the Actual Honeycomb with other candidates, it is necessary to impose some reasonable restrictions on the search space of admissible candidate structures – we are not interested in any kind of surface-minimizing structure, but only in biologically-reasonable ones. Fejes Tóth thus keeps certain structural features of the Actual Honeycomb fixed. These features define a type of mathematical structure called *honeycomb*; this type provides us with a framework, which turns a vague question about the optimal use of wax into a precisely formulated mathematical problem of surface minimization.

The most important features of the type *honeycomb* are that the cells are arranged between two parallel planes, that they are congruent[6], fill the space between the planes without overlap, and that they have an opening in exactly one of the two planes. One consequence of these constraints is that the cells are arranged in two layers, with each cell belonging to exactly one of the two planes, via its opening. We can immediately see that the bee's honeycomb is of the type *honeycomb*.

Fejes Tóth formulates two optimization problems for the *honeycomb*. The *first isoperimetric problem* asks for cells with minimal surface area for a given volume $v$ if we assume that the *honeycomb* has a certain constant width $w$, i.e. a constant distance between the two planes. The effect of assuming a constant width is that we can thereby "fix the depth" of cells: if the width is large compared to the volume of cells, i.e. if $w \gg \sqrt[3]{v}$, then the cells will be stretched orthogonal to the planes. The *second isoperimetric problem* is more permissive in that we let the width of the honeycomb vary as well. Fejes Tóth shows that the bee's solution is not optimal with respect to the second isoperimetric problem, but notes that it might be biologically sensible to let the cells be of a certain, fixed depth. Thus the biologically-relevant problem might be the first isoperimetric problem.

Fejes Tóth then proves the main result of the paper: The bee's cells (figure 4.2) are not the best geometrical solution to both isoperimetric problems. If we choose a different design for the cell bottom, we get cells with a smaller surface area for the same volume and arbitrary lengths; figure 4.3 shows the superior cell.[7] Fejes Tóth sums up his result as follows: "*Instead*

[6] More specifically, only congruent convex polyhedra are considered.

[7] The design of the bee's cell bottom is based on the rhombic dodecahedron, while Fejes Tóth's cells are based on a truncated octahedron.

*of closing the bottom of a cell by three rhombi*, as the bees do, *it is always more efficient to use two hexagons and two rhombi*" (Fejes Tóth, 1964, p. 473, emphasis in original).

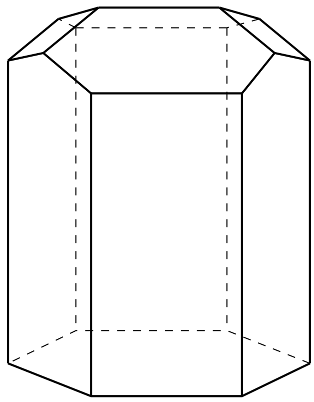


Figure 4.3: The "Fejes Tóth Structure"

The design of this "Fejes Tóth Structure" is superior to the bee's design, but by how much? Fejes Tóth calculates that the saving in surface area is less than 0.35% of the area of a cell opening, and much less of the area of a cell of normal length. He acknowledges that this is a very small gain in efficiency, and as the real cells are not perfectly regular in several respects, it seems that the difference is negligible in practice.

### 4.5.2 Mathematics vs. The World: Discussion of Fejes Tóth

Now that we are familiar with Fejes Tóth's proposal, we want to assess its significance for actual honeycombs. What do the mathematical results tell us about real honeycombs?

Before we tackle this question, it is instructive to compare Fejes Tóth's result with Lyon's and Colyvan's explanation. Fejes Tóth's result differs from Lyon's and Colyvan's in that, if it were applicable, it would establish that the bee's solution is *not* optimal. In this case, it would be interesting to explain, from an empirical point of view, why the bee's solution is not optimal. The second difference between the two results is that Fejes Tóth's is only a relative optimality result. The optimal solutions to both isoperimetric problems are unknown.

Does Fejes Tóth succeed in establishing that the Actual Honeycomb is not optimal? I think not. His solution presupposes that the walls of the mathematical honeycomb are infinitely thin, but neither is this true in the case of actual honeycombs, nor does he provide an argument why the difference between honeycombs with infinitely thin walls and honeycombs with thick walls can be neglected in practice, such that his mathematical results hold in application.

The last point needs some elaboration. In what scenario could we neglect the difference between a honeycomb with infinitely thin walls and a honeycomb with thick walls? The idea behind the assumption of infinitely thin walls could be that all we have to do to find the wall volume of an actual honeycomb, and thus the amount of wax necessary to build it, is to multiply the surface area with a small constant factor that captures wall diameter. We could then compare the wall volume of two de-idealized structures and we would find that the de-idealized Fejes Tóth Structure (cells in figure 4.3 with finite thickness) is better than the de-idealized Actual Honeycomb (cells in figure 4.2 with finite thickness).

However, this procedure will give us a good approximation only if the walls of honeycombs in the world are sufficiently thin and homogeneous, i.e. if the overlap of walls around the cell edges is not too big, if the walls have the same diameter everywhere, and if the structure is regularly shaped. However, this is not the case. Neither are the walls of honeycombs in the world very thin, nor are they homogeneous or particularly regular. This means that if we compare honeycombs in the world with the de-idealized Actual Honeycomb, the difference between the de-idealized Actual Honeycomb and any real honeycomb more than offsets the tiny difference between the de-idealized Fejes Tóth Structure and the de-idealized Actual Honeycomb. The difference found in the mathematical case does not carry over to the real world.[8]

To be fair, Fejes Tóth is clearly aware of the shortcomings of his approach; in the second part of his paper, he takes some steps towards its resolution by taking the thickness of walls into account from the outset. Also, Fejes Tóth's primary goal is to formulate, and solve, a mathematical, not an empirical, problem. He knows that the application of the mathematical result to actual honeycombs is a delicate matter. Consequently, he does not settle for one formulation of the problem, but offers several different approaches.

At this point, we should not be misled into thinking that the problem with the failed explanations we have seen so far is that they use idealizations. The use of idealizations is not problematic *per se*. Lyon's and Colyvan's explanation is a case in point: There are good explanations that describe some three-dimensional setting in two dimensions – think of city street maps. Their explanation could have been successful. The same goes for Fejes Tóth: the assumption of infinitely thin walls might have been fruitful, but it just so happens that the real honeycomb is too irregular for the approximation to be useful.

Nevertheless, Fejes Tóth's proposal means progress: although ultimately not successful, it does not face the problems discussed in the first part of the

[8] Based on this result, it could be conjectured that the small gain in efficiency that Fejes Tóth has found has been "neglected" by the bees because the structure they build is not that regular anyway.

chapter. What is more, Fejes Tóth's discussion suggests how we can fix the problem. We have to take the (irregular) thickness of walls into account. We will examine this kind of approach later in the chapter.

## 4.6 Biological Stories And An Alternative Explanation

In this section, I first review some recent biological examinations of the actual construction process of the bee's honeycomb. We will see that the final verdict on this issue has not been returned. There are two rival accounts: One is that the construction process is, at least partially, physical in that the cells are the result of a liquid equilibrium. The second is that the bees build the cells mechanically.

I then sketch an alternative explanation for the structure of the bee's honeycomb under the assumption that the construction is based on a liquid equilibrium. If this is correct, then the mathematics underlying the explanation is different from what we have seen so far.

This explanation, based on the liquid equilibrium process, serves a double purpose: On the one hand, it is a reasonable proposal for an explanation of the honeycomb. On the other hand, it serves as a further argument against the relevance of the HC and complements the objections I raised in the first part of the chapter. I will present the explanation in a two-dimensional setting to make the argument more accessible; nothing hinges on this in principle.

### 4.6.1 How Honeycombs are Built: Pirk et al.

In their paper "Honeybee combs: construction through a liquid equilibrium process?", Pirk et al. (2004) examine the construction process of the bee's honeycomb and make an observation about the final structure of honeycombs.

First, here is how the bees construct the honeycomb. The bees do not form the hexagonal cells entirely "by hand"; rather, they erect a rough framework of cells which is then melted into its final shape:

> The structure of the combs of honeybees results from wax as a thermoplastic building medium, which softens and hardens as a result of increasing and decreasing temperatures. It flows among an array of transient, close-packed cylinders which are actually the self-heated honeybees themselves. (Pirk et al., 2004, p. 350)

The bees raise their body temperature to over 40°C in a coordinated process – they work simultaneously in neighboring cells in both layers – and the cells get their characteristic hexagonal shape through a thermodynamic process: "the comb structure is a result of a thermoplastic wax reaching a liquid equilibrium" (Ibid., p. 352).

Second, Pirk et al. make an interesting observation. They took a closer look at the shape of the cells, and they found that the cell bottoms, or closings, are not in fact formed by three rhombi (see figure 4.2 for the rhombic design):

> [T]he cell bases are hemispherical from the onset of construction and never form three rhomboids [...] The 'three rhomboids' are just an optical artefact caused by traditional thinking leading to the wrong conclusions about a three-dimensional structure from looking at two interlaced hexagonal rasters in semi-transparency. (Ibid., p. 352)

The apparent three rhombi at the base of cells, the authors contend, is just the structure of the second layer of cells shining through. This finding about the cell bases, as we will see in a moment, has been contested in a recent paper.

If the bees do not construct the cells "by hand", but use a physical mechanism to give the cells their final shape, it is possible to find various optimal solutions not just mathematically, but empirically. In fact, Pirk et al. were able to experimentally confirm the formation of hexagonal cells: they filled the space between rubber cylinders with melted wax and observed the formation of the characteristic hexagonal pattern during the cooling process.

### 4.6.2 The Real Shape of Cell Bases: Hepburn et al.

The results by Pirk et al. have a small but important flaw. The finding that the bottom of cells is spherical, and not formed by three rhombi, has been challenged in a recent paper by Hepburn et al. (2007). Hepburn et al. examined moulds taken from newly constructed, as well as from old, cells and found that while old cells indeed have spherical bases, newly-constructed cells have rhombic bases, see figure 4.4 below. The difference between new and old cells arises "from gradual accretion of silk and larval faeces, which slowly changes the 'apparent' shape of the cell bottom from the real underlying rhomboid to a superimposed hemisphere" (Hepburn et al., 2007, p. 270). Pirk et al. made the mistake of taking moulds exclusively from old cells.

This may appear to be a minor finding, but it is an important part of the bigger picture. Hepburn et al. point out that if the construction mechanism postulated by Prik et al. is correct, the rhombic shape is not that surprising; rather, spherically shaped bottoms would have been "paradoxical":

> [T]hree rhomboids would be expected as a product of equilibrium in precisely the same way that soap bubbles form angular, not hemispherical, contact faces. [...] [T]he underlying geometry

Figure 4.4: mould of new comb, (Hepburn et al., 2007, p. 269)

> of the cell walls and bases are in keeping with mathematical laws on surface minima. (Hepburn et al., 2007, p. 270)

If the cell bases had been spherical, this would have indicated that at least not all parts of cells were constructed by thermal equilibrium; the finding by Hepburn et al. shows that the thermal construction can explain the shape of the cells, provided that we start with the rounded-off cylinders. The real importance of this finding, as well as the connection between the bee's honeycomb, surface minima and soap bubbles, will be further explored in section 4.7.

### 4.6.3 No Liquid Equilibrium After All? Bauer and Bienefeld

A very recent paper, Bauer and Bienefeld (2013), calls into question that the bee's honeycomb is constructed via a liquid equilibrium as postulated in Pirk et al. (2004). Bauer and Bienefeld are much more cautious in their formulations about what we do and do not know about the construction process; they think that the process is still "a mystery".

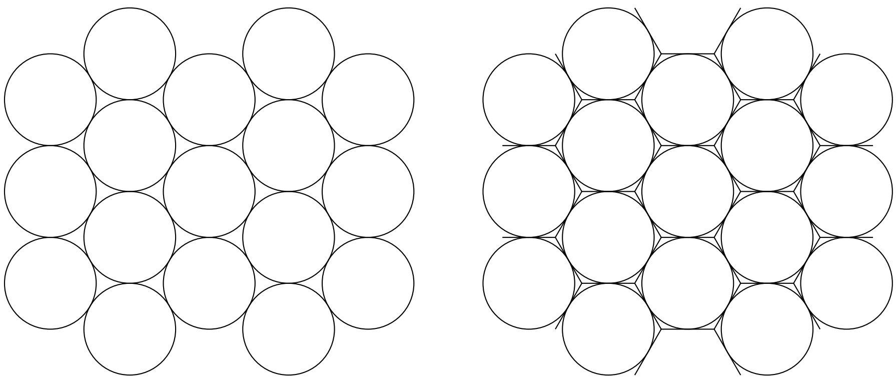
Bauer and Bienefeld examined the construction process of European honeybees (*Apis mellifera*) using infrared and thermographic video observations. Their result is twofold. First, they observe that the bees actively formed the wax with their mandibles, using their legs for stabilization and to apply the necessary force. However, it is unclear how the bees measure the geometrical shape of the cells. Second, the temperature of the wax never rose above 37.8°C, while the thorax temperature of builder bees never rose above 39.6°C, well below the 40°C found by Pirk et al., which is necessary for the liquid equilibrium: "[t]he wax was compact and did not enter the liquid equilibrium state at any of the building temperatures observed" (Bauer and Bienefeld, 2013, p. 48). Bauer and Bienefeld concede that heating up the wax may contribute to its plasticity and thereby may facilitate its (mechanical) shaping.

Finally, they note that wax may not have a clear-cut "jump temperature"; a clear boundary for the phase transition.

### 4.6.4 An Alternative Path to the Hexagonal Grid

The final word on how the bees actually construct the honeycomb is not out yet; both available alternatives raise interesting issues. Here I will sketch how mathematics and biology could be combined to yield a new explanation under the assumption that the the construction process is based on a liquid equilibrium.

The two-dimensional hexagonal grid features in various optimization problems, in addition to the HC; one of them is the circle packing problem.[9] The circle packing problem asks for the optimal packing of equal circles in the plane, i.e. the densest arrangement of circles without overlap. There is an obvious candidate solution: each circle is surrounded by six circles such that the centers of the six circles form a hexagon (see figure 4.5, part (a), below).[10]



(a) (b)

Figure 4.5: Close-Packing of Circles (a), corresponding Voronoi Cells (b)

The optimal packing of circles is called the hexagonal packing, but it is not a tiling of the plane: the circles are not separated by one-dimensional lines, but by a non-negligible, two-dimensional part of the plane (the circles cover only 90.68% of the plane). There is, however, a straightforward and unique way to get from the hexagonal packing of circles to the familiar hexagonal tiling. If we start from the circle packing and divide the remaining space between circles, such that each point is assigned to the closest circle,

[9]See Szpiro (2003, ch. 3-4) for a detailed account of the history of this problem. Lyon and Colyvan mention the result on p. 229, footnote 2. It seem that they find the result to be irrelevant to the honeycomb.

[10]You find this arrangement when you arrange equal coins in the densest possible manner on a table. Although this arrangement clearly is the best solution, a complete proof of this conjecture was only given in 1940 by Fejes Tóth, see Szpiro (2003) for more on this.

we arrive at the hexagonal tiling: the points lying exactly between two or three circles are the points of the hexagonal grid. The cells we get in this way are known as Voronoi cells (see figure 4.5, part (b), for an illustration).

Why should we accept this two-step process as an explanation instead of the HC? Why is this alternative explanation of any relevance at all? After all, the HC at least explains the hexagonal grid straightforwardly – it establishes that a minimum amount of wax is necessary for this structure – while the two-step process seems more complicated: the circle packing problem does not minimize the amount of wax; this only happens in the second step.

The construction process, based on a liquid equilibrium, suggests that the primary optimization problem is a packing problem. In a first step, the bees build cells that make it possible to pack themselves as closely as possible – this is a close-packing of cylinders – bees – in two layers. In a second step, they subtract wax from the close-packed cylinders and thereby minimize the necessary amount of wax. This, I suggest, corresponds to the two-stage process I described above: first, some bodies are close-packed, and then, starting from this close-packing, Voronoi cells are constructed, i.e. as much wax as possible is subtracted. This supports the thesis that the two-dimensional hexagonal tiling is not due to a tiling problem, such as the HC, but to a packing problem, such as the circle-packing problem.

### 4.6.5 What About the Mathematics?

So far, I have proposed that an alternative explanation of the honeycomb's structure is plausible if we assume the liquid equilibrium to be the construction process. On this alternative account, the real optimization problem is not surface minimization *tout court*, but a packing problem with subsequent surface minimization. It could be asked what mathematics has to say on this issue: What are the optimal solutions to the packing and surface minimization problems in three dimensions? And; How are the two problems related?

Unfortunately, the answers to these questions are not known. For example, it is an open problem as to whether the obvious arrangement of parallel rows of congruent cylinders is the best solution to the general "tin can stacking problem" in three dimensions; see Hales (2000, p. 440). It is also unknown whether the bee's arrangement of cylinders in two layers is optimal. The same is true for the relation between the different optimization problems. This is a common phenomenon, as I pointed out above.

In conclusion, based on one possible construction process, the primary optimization problem seems to be a three-dimensional packing problem, not a three-dimensional tiling problem. Therefore, if a two-dimensional optimization problem is relevant to the shape of the openings, it it is plausible that this is the circle-packing problem, not the HC. However, it is far from conclusive that the alternative proposal will carry the day. The situation is just not well enough understood from a mathematical, as well as from a biological, point of view.

## 4.7 Dry Foams, Wet Foams, Honeycombs

At the end of the last section, I pointed out that we do have a very incomplete understanding of three-dimensional optimization problems: We do not know what the solutions to relevant packing and surface minimization problems are. However, there is a physical-mathematical framework that unifies all the problems we have seen so far. This framework is the topic of the present section.

The framework suggests an alternative path of inquiry: While we do not yet understand the mathematics of three-dimensional honeycombs, we can approach the problem experimentally. Two physicists have explored this option: they produced a physical realization of the bee's honeycomb. The result of the experiment reveals a deep connection between close-packing and surface minimization problems, bringing together all mathematical approaches to the bee's honeycomb discussed above.

### 4.7.1 A Physical-Mathematical Framework: Foams

The distinction between optimal structures with finite and infinitely-thin walls, or between packing and surface minimization problems, can be cast in terms of foams.[11] Why foams? Just as it is an evolutionary advantage for the bees to minimize the amount of wax when building a honeycomb, it is energetically advantageous for soap bubbles to minimize surface area. The distinction between optimal cellular structures with infinitely-thin walls and cellular structures with thick walls carries directly over to foam structures: Some foams have negligible liquid content and the boundaries between soap bubbles are (almost) infinitely thin. These are called *dry foams.* On the other hand, foams with a non-negligible liquid content are called *wet foams.* In wet foams, the bubbles do not have large common boundaries, but minimize surface energy one by one. This results in spheres floating in some medium.

We can classify all optimization problems we discussed so far into one of these two kinds of foams. In two dimensions, the HC is about the optimal dry foam of unit bubbles, while the circle-packing problem is the analogue wet-foam problem. In three dimensions, there is the sphere packing problem, or Kepler problem, which is a wet-foam of unit bubbles (spheres); the analogue dry-foam problem, also known as the Kelvin problem, asks for the optimal dry-foam made of bubbles, or cells, of unit volume.[12]

[11] My account in this section is based on Klarreich (2000) and Hales (2000). Weaire and Hutzler (1999) is an introduction to the physical aspects of foams.

[12] Today three of these four problems are solved; only the Kelvin problem remains open. Mathematicians believe that it could take decades to solve it. For a long time,

The foam problem, corresponding to the honeycomb, cannot be one of these four problems. The honeycomb foam has to be bounded because every cell has to have an opening. We have already encountered dry-foam problems that seem to capture the bee's honeycomb very well: the two isoperimetric problems based on *honeycomb*-type structures defined by Fejes Tóth. The main shortcoming of these formulations is the assumption of infinitely-thin walls, in other words, that it is a dry-foam problem. What we would have to examine, then, is the wet-foam analogue of Fejes Tóth's problems: we have to add some liquid content to *honeycomb*-type structures.

I pointed out above that we do not know whether the Fejes Tóth Structure is the solution to the dry-foam version of the *honeycomb*, and the solution to the corresponding wet-foam problem is not known either. What can be done, and has been done, is to carry out experiments with foams. We will now turn to such an experiment.

### 4.7.2 Realizing the Honeycomb, Physically: Weaire and Phelan

Denis Weaire and Roberd Phelan report their experiment in a paper entitled "Optimal design of honeycombs". They start by summing up the results by Fejes Tóth; see section 4.5.1 above. Then they describe their experiment, which reproduces the *honeycomb* with soap bubbles. To this end, they introduce equal-sized bubbles of a liquid solution, between two glass plates, such that they form a double layer of cells. This is a bounded foam. Weaire and Phelan observe that the two layers form an array of cells with hexagonal openings. Their main finding is that, as the liquid content of the foam is varied, the design of the cell bottoms changes:

> [In the case of a] foam of low liquid content [...] which corresponds to Toth's geometrical picture, we observe the structure proposed by him [see figure 4.3]. If we wet the foam by the addition of more liquid, the junction between films are thickened to form what are called Plateau borders. The conditions under which surface energy is to be minimized are accordingly changed. What happens as the foam is progressively wetted is quite dramatic. At a certain point the Toth structure becomes unstable and there is a sudden switch to the configuration favoured by the bees [see figure 4.2]. Such a switch also takes place in the reverse direction, as liquid is removed. (Weaire and Phelan, 1994)

The experiment, then, has three main components. First, the Fejes Tóth Structure is realized as a real dry foam. Second, the Actual Honeycomb

---

the candidate solution was a structure proposed by Kelvin, but in 1994, Denis Weaire and Robert Phelan found a counterexample to Kelvin's structure, now called Weaire-Phelan structure.

is realized as a wet foam. Third, and most surprisingly, there is a real, physical connection between these two realizations in that one structure is transformed into the other if liquid content is added or subtracted. Let us now turn to the interpretation of this result.

### 4.7.3 Weaire and Phelan: Interpretation

We can interpret Weaire's and Phelan's finding as a sort of physical verification that both the Fejes Tóth Structure (figure 4.3) and the Actual Honeycomb (figure 4.2) are optimal solutions: the Fejes Tóth Structure is the optimal solution to the dry-foam version of the problem, while the Actual Honeycomb is the optimal solution to the wet-foam version. Of course, this experiment in no way constitutes a mathematical answer to the mathematical dry-foam and the wet-foam problems. A proof of optimality is in both cases missing.

We have seen in the previous section that according to one possible construction process, the bees construct the honeycomb using a liquid equilibrium. They melt the cells of the honeycomb into their final form. This means that the bees realize actual foam bubbles in the honeycomb: the wax becomes liquid and obeys the physical laws of a foam. The bee's honeycomb is an instance of an actual wet foam: the underlying physical process is the same in the case of the bee's honeycomb and in the experiment. Therefore, it is no surprise that they both realize the same structure.[13]

What does the fact that we can physically transform one foam into the other, by adding and subtracting liquid content, mean for the bees honeycomb? We could speculate that the bees start with a wet foam and melt away some of the wax. This could correspond to the process involving Voronoi cells I described in the previous section. However, if the bees continued to melt away wax, it could happen that a sort of structural transition from the Actual Honeycomb to the Fejes Tóth Structure occurs. If this is so, then Fejes Tóth's proposal does not apply simply because the bees, for whatever reason, do not push the melting to the extreme. Ultimately, we would like to understand the mathematics behind this structural transition.

## 4.8 Conclusion and Outlook

Here is a summary of the most important results and lessons of the chapter.

[13]Hypotheses linking the bee's honeycomb to soap bubbles via an equilibrium process have been around for some time; see e.g. Klarreich (2000, p. 159). However, only the biological results reported in the last section confirm that this could actually be the case. Note that the biological results were unknown to Weaire and Phelan in 1994! Still, one important difference between a genuinely liquid foam and the bee's honeycomb should be kept in mind: the bee's foam is not liquid in its totality, but only locally. It is unclear to me how this affects the formation of minimal surfaces.

The first goal of the chapter was to argue against the original explanation of the bee's honeycomb, proposed by Lyon and Colyvan. I offered two arguments. First, I established that the original explanation is incomplete because the real honeycomb is non-trivially three-dimensional. Second, I argued that, because the real structure is three-dimensional, the HC could be superfluous, given that it is a three-dimensional optimization problem.

The main philosophical lesson we can learn from the first part of the chapter is that we have to be more careful in the use of examples from science, especially if we rely on our examples to be real-life cases. The discussion also has immediate consequences for the debate on mathematical explanations in science. As the original explanation, based on the HC, is not scientifically adequate, we should stop using it as a case of mathematical explanations in science and postpone the discussion of the role of mathematics for the structure of the bee's honeycomb until we have a clear, well-founded explanation of this phenomenon. As a case in point, my argument undermines the use of the honeycomb case as a counterexample by Alan Baker.

The second goal was to present and assess mathematical and scientific results that could lead to a novel, adequate explanation. In the discussion of Fejes Tóth's mathematical proposal, I found that his result is not applicable to the bee's honeycomb because it assumes infinitely-thin walls, an idealization that cannot be removed in a sensible way. I then presented an overview of some relevant biological results. Here I found that the biological details of the construction process are still debated. There are two rival accounts. In the first account, the construction process relies on a liquid equilibrium, while the second account maintains that the construction is carried out mechanically by the bees.

If the first construction process, using a liquid equilibrium, is correct, then the best close-packing of cylinders, followed by a subtraction process from the close-packing structure, may be relevant instead of a tiling problem. This yields a different mathematical-biological explanation of the structure of the bee's honeycomb. However, the relevant mathematical optimization problems are still open.

It is not clear to me what kind of mathematical optimization problem is relevant if the second, mechanical construction process is correct. This outcome would raise interesting biological questions as well as issues about the application of mathematics. Bauer and Bienefeld point out that it is not clear how the bees measure the geometry of the cell. It is even more puzzling how the bees solve a mathematical problem, which seems necessary if they form the cells by hand. How do they memorize the solution to this optimization problem? This question becomes even more pressing if the underlying mathematics is very complicated. How do the bees bypass the complicated mathematics, and find an optimal solution in a potentially very large search space? If the number of alternatives is infinite, it seems impossible for them to check all possible alternatives by "trial and error"; the bees would have to perform a kind of intelligent search. How is this supposed to work?

Irrespective of what the correct biological account of the construction process is, we see that biological results, mathematical considerations, and even physics are intertwined. The actual construction process can speak in favor of a particular mathematical account of the story, but it can also make it implausible. For example, if the construction relies on a liquid equilibrium, it is plausible that some wet foam formulation of the underlying mathematics is in action, while a mechanical construction process could rule out the application of this kind of mathematics, as it could be impossible for the bees to realize a very complicated mathematical structure "by hand". The physical realization of the bee's honeycomb as a wet foam speaks in favor of the construction based on the liquid equilibrium.

Then, we saw once more that the issues of application of mathematics and idealization are intimately connected. Many of the problems of application we discussed can be cast in terms of harmless and harmful idealizations. In the present case, as in general, we have to make an effort to explain why some idealizing explanation fails. We glossed over other idealizations that are more or less unproblematic; for example, we always assumed that the structures are infinitely extended in two dimensions – but that is no problem as long as the real honeycomb is sufficiently large. To say that an explanation using mathematics fails because mathematical models idealize, i.e. contain information that is literally false, is always an easy way out, but we should resist this temptation. A further analysis of this case, in terms of idealizations, would certainly be fruitful.

Finally, there are some open scientific and mathematical problems. First, the explanations we discussed so far only take the (biological) desideratum of optimization into account. However, this cannot be the only constraint at play. For example, the bees have to make sure that the honeycomb is sufficiently stable.[14] The actual honeycomb might not reach the point of structural transition realized in the Weaire-Phelan experiment because stability stands in the way. Stability is obviously in tension with optimization – but how exactly does stability come into play and how do the constraints interact?

Then, I did not discuss all relevant kinds of honeycombs in this chapter. Different kinds of bees might construct honeycombs with a different global structure. This would open the possibility of thinking about other kinds of mathematical *honeycombs*, and their respective advantages and drawbacks.

I will not repeat the many open mathematical problems I pointed out throughout the chapter. However, I would like to stress that it is not sufficient to study a phenomenon such as the bee's honeycomb exclusively from a mathematical point of view. It is a job for philosophers of science to try to

[14] Weaire and Hutzler (1999, p. 167) mention stability and simplicity as constraints besides surface minimization.

connect disparate results, from various disciplines, to get the bigger picture; be it only to understand the bee's honeycomb.

# Part II

# The Application of Mathematics in the Genesis of General Relativity

# Chapter 5

# The Genesis of General Relativity: A Short Introduction

## 5.1 Introduction

This chapter is a short introduction to the history of early general relativity, between 1907 and 1916. The goals of the chapter are, first, to provide the historical background for the philosophical discussion in chapter 8, and second, to give a short, informal introduction to the physics and mathematics relevant to the episode.

The chapter is organized as follows. In section 5.2, I introduce the central motives, theories and strategies of Einstein's search for GR. After motivating the search for GR in subsection 5.2.1, I introduce relevant physical concepts and theories that were already available at the time, in subsection 5.2.2. In subsection 5.2.3, I discuss the history of the "new mathematics", what is now known as tensor calculus, the mathematical theory commonly used to formulate GR. In subsection 5.2.4, I introduce heuristic strategies and principles that guided Einstein in his search for GR.

These protagonists and motives then enact the drama of GR[1], in sections 5.3, 5.4 and 5.5: The first two acts of the drama cover the episode from 1907 to 1912, the year in which Einstein and Grossmann formulated the so-called "Entwurf" theory of GR. The third act tells the story of how the discovery of the final, correct field equations was delayed for three years, despite the fact that all main protagonists were already on the scene.

The account presented here does not pretend to be original; it is largely based on the standard reference for the history of GR, the four volumes of "The Genesis of General Relativity", in particular Janssen et al. (2007a,b);

[1] The picture of GR as a drama goes back to Stachel (2007).

Renn and Schemmel (2007). The discussion of the purely mathematical tradition in 5.2.3 draws on the historical account in chapter 6, and the historical literature therein.

## 5.2 Protagonists and Motives

### 5.2.1 What prompted the search for GR?

One of the starting points for the search for a generalized theory of relativity (GR) was a conflict that arose between classical mechanics, in particular Newton's universal law of gravitation, and the new theory of special relativity (SR), discovered by Einstein in 1905.

To a large degree, SR grew out of the then-new theory of electrodynamics. SR is based on two principles. First, the principle of relativity states, that if a physical law holds in one inertial frame, it holds in all inertial frames. Second, the principle of the constancy of the speed of light states that the speed of light, $c$, is the same in all inertial frames. The second principle is often read as the claim that the speed of light is an upper limit on the speed of physical interactions. These two principles together imply that the laws of physics are Lorentz covariant, i.e. they do not change their form under so-called Lorentz transformations.[2]

Einstein's main achievement was to reinterpret the formal Lorentz transformations, which were known to hold in electrodynamics, as a fundamental feature of space and time; it is particularly noteworthy that in SR, the fact that physical interactions propagate, at most, at the speed of light becomes a feature of space-time, and thus is an explicit desideratum for all space-time theories.

The requirement that physical interactions travel at a finite speed is in contradiction to Newton's law of gravitation, which implies an action-at-a-distance between masses. It was thus necessary to reformulate the law of gravitation, to make it Lorentz covariant. On top of this, it was desirable to implement one of the features of electrodynamics in gravitational theory, namely to find a field theoretic formulation of gravitation, such that forces are mediated by a real gravitational field. Finally, SR also had a deep impact on the concepts of mass and energy. It showed that they are equivalent; if mass is the source of gravitation, then so is energy. A new law of gravitation had to incorporate the equivalence of mass and energy. All these points made it necessary to rework gravitation in view of SR.[3]

[2]See Renn (2007a, sec. 2.8) for a short historical introduction to the genesis of SR.

[3]See Renn (2007a, sec. 2.9) for more on the conflict between SR and classical gravitation.

### 5.2.2 Physics before GR

What is the role of existing physical theories in the genesis of GR? In particular: which theories were important for the formulation of GR, and in what form? In this section, we introduce important classical and special-relativistic theories and concepts.

#### Minkowski Formalism (SR)

Special relativity is a necessary condition for the formulation of a generalized theory of relativity. One formulation of SR, Minkowski's geometrical approach, was particularly important in the genesis of GR.

Einstein's formulation of SR, in 1905, had shown that space and time themselves were no longer concepts that could be treated separately – they were no longer invariant quantities. Minkowski's reformulation of SR, in 1908, established that the two principles of SR imply the existence of a different invariant quantity, now known as the Minkowski line element, which, in turn, can be used to succinctly describe special-relativistic space-time.[4]

The significance of the Minkowski line element is best illustrated with an analogy. In Cartesian coordinates[5], we can express the distance $ds$ between any two points with coordinate differences $dx, dy, dz$[6] in three-dimensional space with the help of the Pythagorean Theorem:

$$ds^2 = dx^2 + dy^2 + dz^2 \tag{5.1}$$

This form of $ds^2$ is an invariant quantity under change of Cartesian coordinates: If we use any other set of Cartesian coordinates, $ds^2$ is the same. On the Cartesian picture, length is not a property of the coordinate system, but a real geometric property. Minkowski showed that the same can be done in SR. If we multiply the time coordinate with the speed of light, and add it to the Pythagorean formula with negative sign, we get a metric with four spatial coordinates:

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2 \tag{5.2}$$

$ds$ is the Minkowski line element; it represents lengths not in space, but in space-time. It is sometimes called a pseudo-Euclidean metric. It is similar to Pythagorean distance, but has some peculiarities. $ds^2$ can be zero without the components being zero, namely if $dx^2 + dy^2 + dz^2 = c^2dt^2$. Two events that are separated in this manner are connected by a light signal. The set

[4]See e.g. Renn (2007a, sec. 2.8) for an overview of the historical significance of Minkowski's work.

[5]Cartesian coordinate systems are orthogonal with equal units for all coordinate axes.

[6]Note that historically, the necessity of using coordinate differentials was not a feature of Minkowski's original formulation, but only emerged in the so-called Einstein-Abraham controversy, see below.

of all events, connected by a light signal with $e$, form the light cone of $e$. If the distance between two events is negative, i.e. $c^2dt^2 > dx^2 + dy^2 + dz^2$, they are time-like separated, i.e. one is inside the other's light cone. If the distance is positive, they are space-like separated, i.e. they lay outside each other's light cone. This partition of space-time, into three invariant sets, is best captured by space-time diagrams.

In four-dimensional geometry, Lorentz transformations are analogous to spatial rotations in three dimensions. It can be shown that $ds^2$ is in close correspondence with the Lorentz transformations: it is invariant under the Lorentz transformations, and it can be used to characterize them. $ds^2$ is independent of coordinates on the picture of SR.[7] In sum, Minkowski provided a geometric formulation of SR, as a theory of a four-dimensional, pseudo-Euclidean space-time manifold.

Minkowski's reformulation is significant, for the genesis of GR, for several reasons. It is one of two crucial ingredients that paved the way for a new mathematical approach; the generally covariant formulation of relativity, see Norton (1984, p. 260). Firstly, it suggested a characterization of physically significant quantities in geometric and invariant theoretic terms. Secondly, as we will see, the form of the expression (5.2) probably suggested a generalization of the metric. It took some time before Einstein recognized the importance of Minkowski's formulation of SR; he dismissed it as a fancy piece of mathematics probably as late as 1911; see Norton (2000, p. 141). In 1912, it became an important part of his heuristics.

### Lorentz Model (Electrodynamics)

We noted above that SR at least partially grew out of Einstein's reflections on certain features of electrodynamics. Electrodynamics continued to be a template for Einstein's approach to an integration of gravitation and relativity. The approach has been characterized as the "Lorentz model" of physical theories.[8]

The Lorentz model grew out of Hendrik Antoon Lorentz's reflections on experiments in electrodynamics. On the Lorentz model, the interaction between elementary particles should not be described in terms of a force that the particles exert on each other, instead we should think of the interactions as mediated by a field created by the particles. Only when this field given can we determine how particles move, as it is the field that exerts the force on particles.[9]

---

[7] Note that Minkowski did not introduce $ds$ as "line element"; this only happened in discussions of rigid bodies and rigid motion; see Stachel (2007, p. 105).

[8] See Renn (2007a, p. 54) for a useful, short description of the Lorentz model. For a more detailed explanation see Renn and Sauer (2007, sec. 2). The following draws on these accounts. Note that I do not take this model to carry any metaphysical weight.

[9] In principle, particles can be manipulated independently of a field, e.g. with other kinds of forces, which, in turn, influence the fields.

This model of a physical theory suggests a division of labour in the mathematical representation: First, we have to describe how the distribution of elementary particles determines a corresponding field with a field equation – in electrodynamics, these are the Maxwell equations. Second, can then deduce an equation of motion from the field equation, which describes how a test particle moves in the field; this is the Lorentz force in electrodynamics. In classical mechanics, the Poisson equation is the field equation from which we can derive the Newton's law of motion.

**Poisson Equation (Classical Mechanics)**

Two aspects of classical mechanics stand out as particularly important for the discovery of GR: the formulation of classical gravitational theory with the Poisson equation, and Lagrangian mechanics.

The Poisson equation is one of the starting points of Einstein's search for the field equation of GR.[10] It encodes the classical theory of gravitation, but in a form that was particularly suitable for Einstein's purposes. The Poisson equation is

$$\Delta\phi = 4\pi G\rho \tag{5.3}$$

The components of the equation are the gravitational potential $\phi$, the Laplace operator $\Delta$, defined as

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{5.4}$$

Then, $G$ is the gravitational constant, and $\rho$, is the density of gravitational matter. Both $\phi$ and $\rho$ are scalar functions of three-dimensional space.

Why is the Poisson equation a more suitable starting point for the search of GR than, say, Newton's law of gravitation? The latter is

$$\mathbf{F} = G\frac{mM}{r^2}\mathbf{e}_r \tag{5.5}$$

with $\mathbf{F}$ the force vector, $m$ and $M$ two point masses, $\mathbf{e}_r$ the unit vector in the direction of the force, and $r$ the distance between the two masses.

The differences between the two formulations of classical gravitation can help us get a clearer picture of the role of the Poisson equation. First, Newton's equation is a force law; it is a vectorial equation. The Poisson equation is a scalar potential equation. We can recover the gravitational force from the the potential, determined by the Poisson equation by differentiation, but the fact that the Poisson equation is scalar makes it more tractable in many cases. This is a pragmatic difference between the two formulations.

[10]The following account, of the role of the Poisson equation in the genesis of GR, is based on Renn and Sauer (2007, sec. 2.1).

A second difference is that Newton's law is formulated for two point masses, while the Poisson equation uses a continuous distribution of mass densities. Thus the Poisson equation is more general. Despite these differences, the two formulations are often characterized as mere formal variants with the same physical content.

The crucial difference between the two formulations is that the Poisson equation is closer to the Lorentz model: The potential can be interpreted as a gravitational field, which is determined by the distribution of mass density. From the potential we can derive the force law. This suggests that the Poisson equation is a good starting point for a reformulation of gravitational theory, based on the Lorentz model.

The analogy between the Poisson equation and an actual field theory is incomplete: in the domain of classical mechanics, the potential function is not interpreted realistically – it is a mathematical tool. The advent of GR will put the gravitational potential right at the center of attention; this is a radical conceptual shift.

### Energy-Momentum Tensor (SR, Continuum Mechanics)

The Energy-Momentum (EM) tensor is the mathematical object that represents the unification of mass, energy, momentum and stresses in GR. However, we do not need GR to construct the EM tensor; it has its roots in the special-relativistic reformulation of branches of classical mechanics, such as hydrodynamics and electrodynamics. As soon as SR was available in a vector-analytical formulation, it could be inferred that the EM tensor should replace the simpler, classical concept of energy and momentum. Here we will explain, very briefly, the motivation for the shift from the simple classical concepts to the rather involved EM tensor.[11]

The EM tensor describes continuous distributions (densities) of quantities – energy and momentum in the case of GR. The densities are assigned to small elements that fill space-time. The picture that is often adopted is that of a fluid. A fluid is different from a solid in having small antislipping forces: The forces that counteract motion in the direction parallel to the surface between two neighboring elements are small. In a perfect fluid, these forces are zero. There is an even simpler case than that of a perfect fluid, that of "dust". In dust, particles in an element are not in relative motion – there is an inertial frame in which all particles in the element are at rest. This simple case is sufficient to understand why we need a rank-two tensor to represent mass-energy and momentum in SR, and consequently GR.

The first ingredient for the argument is the reformulation of the concept of density in SR. Consider an element of dust, in the inertial frame, in which

[11] The account given here is based on the very accessible presentation in Schutz (2009, ch. 4) – as a consequence, the account given here is not entirely historically accurate. I will indicate which parts of the account can be found in Einstein's writings.

the particles are at rest. The number of particles $n$ in an element is called the number density of the element. The number density is not a Lorentz covariant quantity: if we consider the same volume element in a different inertial frame, the volume will contain the same number of particles, but it will be Lorentz contracted in the direction of motion, which increases the number density. However, we can integrate the number density as a component in a Lorentz covariant four-vector, the number-flux four-vector. The other components of this vector capture the number of particles flowing through a unit surface of constant $x, y, z$ in a unit of time. It makes intuitive sense that this is a Lorentz covariant four-vector, if we conceive of the time component of the vector, the number density, as the flux of particles through a surface of constant time – flux in time and space are on a par in SR.

The second ingredient is very similar, but concerns the concepts of energy and momentum in SR. In isolation, they are not Lorentz covariant, but they can also be integrated as the components of the energy-momentum four-vector, the four-velocity of a particle multiplied by its mass. Energy (or mass) are the time-component of this vector, while the three components of three-momentum are the spatial components.[12]

These two ingredients imply that we need a second-rank EM tensor to represent the distribution of mass-energy and momentum in space-time. For dust, we can define the energy density of a volume element as the energy of each particle in the rest frame, $m$, times the number of particles in the volume element, $n$, yielding the energy density $\rho = mn$. If we now look at how energy density transforms under change of inertial frame, we see that, according to the first ingredient, we have to transform it, because the volume element is Lorentz contracted, and according to the second ingredient, we have to transform it using a second Lorentz factor, because energy transforms into momentum. Thus, energy density transforms, not as the component of a four-vector, but as the component of a quadratic array with sixteen components, a rank-two tensor that is the tensor product of the number-density vector and the energy-momentum four-vector. The same argument can be applied to the other components of the EM tensor $T^{ab}$, the energy flux as well as the momentum density and flux.

The case of dust is also discussed in Einstein and Grossmann (1995, par. 4) under the title "motion of continuously distributed incoherent masses in an arbitrary gravitational field". Einstein writes the EM tensor as

$$\Theta_{\mu\nu} = \rho_0 \frac{dx_\mu}{ds}\frac{dx_\nu}{ds} \tag{5.6}$$

This is exactly what is described above, with $\rho_0$ the mass-energy density and $\frac{dx_\mu}{ds}\frac{dx_\nu}{ds}$ the product of the two vectors described above. Note that we have adopted Einstein's notation, which does not use the position of indices

[12] See Schutz (2009, ch. 2).

to distinguish co- and contravariant quantities – the contravariant nature of the EM tensor is indicated by the use of the Greek letter Θ.

As Einstein notes (Einstein and Grossmann, 1995, p. 312), the EM tensor will take on a different form if we consider more general situations, such as perfect fluids. In general, the components of $T^{ab}$ will be the flux of $a$ momentum across the surface of constant $b$ in some frame of reference, with "time momentum" being energy, and flux across a surface of constant time the number density.

### Lagrangian Formalism (Classical Mechanics)

The Lagrange formalism of classical mechanics can be used to solve mechanical problems under geometrical constraints.[13] On the Newtonian approach, geometric constraints have to be formulated using constraining forces, which complicates matters. The Lagrangian formalism does not use forces, but formulates mechanical problems in terms of energy, and as constraining forces do not do work, they are easier to capture in this framework.

More specifically, the Lagrangian of a physical system is the difference between kinetic and potential energy of the system:

$$L = T - V \tag{5.7}$$

The idea is that the equation of motion can be obtained by considering all possible paths of a body, where the paths under consideration already incorporate geometrical constraints. The actual path is found by calculating the extremal integral of $L$. The principle warranting this inference is called Hamilton's principle, and it can be expressed as

$$\delta\Big\{\int L dt\Big\} = 0 \tag{5.8}$$

From this principle, we can deduce the equation of motion, the so-called Euler-Lagrange equations:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_i}\right) - \frac{\partial L}{\partial x_i} = 0 \tag{5.9}$$

The Euler-Lagrange equations are analogous to Newton's law of motion: $\frac{\partial L}{\partial \dot{x}_i}$ can be read as the impulse, and $\frac{\partial L}{\partial x_i}$ as the force; we thus obtain the result that the change of impulse over time equals force. The use of the Lagrangian formalism constitutes a shift from a formulation of mechanical problems in terms of forces to a geometrical formulation.

Later in the genesis of GR, the Lagrangian formalism also suggested interpreting the metric as the gravitational potential, for the following reason. We can think of a geodesic line, the shortest distance between two

[13] The following account is based on Renn and Sauer (2007, pp. 141, 155).

points in a general space – informally the closest thing to a straight line in a curved surface, think of big circles on a globe – as the extremal integral of an infinitesimal line element $ds$ (which is already a generalization of the Minkowski line element in equation (5.2)):

$$\delta\left\{\int ds\right\} = 0 \tag{5.10}$$

There is now an analogy between this purely geometrical principle and Hamilton's principle in equation (5.8), suggesting that we use $ds = Ldt$. If we insert this into the Euler-Lagrange equations, we see that the force term $\frac{\partial L}{\partial x_i}$ is basically a (coordinate) derivative of the metric, i.e. the metric has exactly the same role as the potential in classical mechanics. This leads to an identification of the metric with the potential.

### 5.2.3 Enter the Mathematics

In this section, we give a very short overview of the "new" mathematics that was first applied in the Entwurf theory.

#### Gaussian Surface Theory

Gaussian surface theory, a part of modern differential geometry, is the mathematical theory that purportedly established a close connection between geometry and GR from early on. In later recollections, Einstein repeatedly wrote that Gaussian surface theory helped pave the way for the search for a suitable mathematical framework for GR; see Reich (1994, ch. 5) and Stachel (2007). Here we give a very short, informal overview of the theory.[14]

Intuitively, a (regular) surface is a two-dimensional landscape in ordinary, three-dimensional space, without edges and self-intersections. Examples of surfaces are planes, spheres and rotational paraboloids. Two-dimensional surfaces are parametrized by a mapping from two-dimensional coordinate space to real three-dimensional space; the mapping is a local description of the surface and ensures there is a (unique) tangent plane to every point of the surface.

For our purposes, Gauss's most important result was to show that certain quantities of surfaces can be described intrinsically, i.e. without using the embedding space of the surface. The central object of intrinsic geometry is the so-called first fundamental form. The first fundamental form is a measure of infinitesimal distances in a Gaussian surface. It can be expressed in different ways. If we use the three-dimensional ambient space, an infinitesimal displacement from some point in direction $w = (x, y, z)$ of the surface is

[14] The following account is based on the historical accounts just mentioned, and on do Carmo (1993), an introduction to differential geometry.

$$dw^2 = dx^2 + dy^2 + dz^2 \tag{5.11}$$

This is a form of the Pythagorean theorem. Gauss's innovation is based on the fact that we can express the first fundamental form in terms of the coordinate function of the surface in that point. Take a vector $v = (p, q)$ in two-dimensional coordinate space, and map it to a three-dimensional vector tangent to the surface, using the parametrization of the surface: $w = (x(p, q), y(p, q), z(p, q))$. This results in a description of an infinitesimal displacement in terms of the two-dimensional coordinate space:

$$dw^2 = Edp^2 + 2Fdpdq + Gdq^2 \tag{5.12}$$

The letters $E$, $F$ and $G$ stand for the following functions:

$$\begin{aligned} E = &\ \left(\tfrac{dx}{dp}\right)^2 + \left(\tfrac{dy}{dp}\right)^2 + \left(\tfrac{dz}{dp}\right)^2 \\ F = &\ \tfrac{dx}{dp}\tfrac{dx}{dq} + \tfrac{dy}{dp}\tfrac{dy}{dq} + \tfrac{dz}{dp}\tfrac{dz}{dq} \\ G = &\ \left(\tfrac{dx}{dq}\right)^2 + \left(\tfrac{dy}{dq}\right)^2 + \left(\tfrac{dz}{dq}\right)^2 \end{aligned} \tag{5.13}$$

$E$, $F$ and $G$ encode the rate of change of a tangent vector $w$, in terms of the rate of change of a coordinate vector $v = (p, q)$, in coordinate direction $p$ (E) and $q$ (G), and the degree of dependence of the coordinate directions $p$ and $q$ (F). The significance of the first fundamental form with components $E$, $F$ and $G$ lies in the fact that it allows us to determine intrinsic geometrical properties of a surface, such as the length of curves in the surface, angles between two curves, surfaces areas, and intrinsic curvature. One of the most important results about intrinsic properties is Gauss's "theorema egregium", which establishes that the intrinsic curvature of a surface depends only on the first fundamental form, and not on the parametrization of the surface.

In modern terminology, the functions $E, F, G$, are the components of the metric of the surface. If we can map one surface to another, and preserve these functions, the two are (locally) isometric, which means that "from within the surface", there is no way of telling the two surfaces apart.

Gauss also showed that the shortest curves within a given surface depend only on the metric of the surface. We can parametrize a curve in coordinate space as $c(t) = (p(t), q(t))$; this can be plugged into a parametrization of the surface to get the curve embedded in the surface in three-dimensional space. The shortest distance between two points can now be found by finding a minimal curve between the two. The distance between two neighboring points is given by

$$ds = \sqrt{Edp^2 + 2Fdpdq + Gdq^2} \tag{5.14}$$

Extremal paths between $a$ and $b$ can therefore be found as

$$\delta\left\{\int_a^b \sqrt{Edp^2 + 2Fdpdq + Gdq^2}\right\} = 0 \tag{5.15}$$

This equation can be found, more or less in this form, in notes that Marcel Grossmann took from lectures on differential geometry by Carl Friedrich Geiser – and these lecture notes, we are told, were used by Einstein to learn for the exams.[15]

### Riemann: Manifold, Metric, Curvature Generalized

A further seminal contribution to differential geometry was provided by Riemann (1876c), entitled "Ueber die Hypothesen, welche der Geometrie zu Grunde liegen".[16] This is the written version of Riemann's habilitation lecture. It addresses a wide audience and therefore is more "philosophical", or conceptual, rather than mathematical, in that it contains almost no formulas.

The ideas laid out in the habilitation lecture generalize Gauss's ideas to general, $n$-dimensional manifolds, which need not be embedded in a space of higher dimension. The habilitation lecture begins with the first ever discussion of general manifolds, including their topological and geometrical properties.[17]

Riemann then discusses what we would call a Riemannian metric, i.e. a quantity $ds$ which is a function of variables $dx_i$ with $n \cdot \frac{n+1}{2}$ coefficients. While he does not write down the expression, he is referring to the line element and the components of the metric tensor $g_{ik}$ as coefficients of variable differentials:

$$ds^2 = g_{ik}dx_i dx_k \tag{5.16}$$

Riemann notes the tensorial character of the line element: it is independent of the choice of variables. The line element, being an "inner measure" ("inneres Massverhältnis") of a manifold, is independent of points outside the manifold, i.e. of its embedding. This suggests a generalization of intrinsic geometry to $n$ dimensions.

The habilitation lecture also contains the first description of the "curvature measure" ("Krümmungsmass"), what we know as the Riemann tensor. The metric has $n \cdot \frac{n+1}{2}$ components. Only $n$ of these are determined by changes of the $n$ variables; $n \cdot \frac{n-1}{2}$ are independent. According to Riemann, these are determined by the "curvature measure", or by the nature of the

[15] See Stachel (2007, p. 104) for the expression of the geodesic, and Reich (1994, sec. 5.1) for more on Einstein's mathematical education. See Sauer (2013) for evidence that Einstein actually may have written a brief note ("Krakeleien") in Grossmann's lecture notes.

[16] The following account draws on Reich (1994, sec. 2.1.3.1.).

[17] See Scholz (1980) for more on Riemann's notion of manifold.

manifold. If the "curvature measure" is zero, the manifold is flat. The notion "curvature measure" is due to Gauss; Riemann's more general notion coincides with Gaussian curvature, up to a constant, for surfaces.

Riemann returned to the Riemann curvature tensor in Riemann (1876a), the "Commentatio", a paper that appeared posthumously. One section of this paper is about the transformation behavior of the metric; more specifically, it asks under which conditions the metric vanishes under variable transformations. Riemann shows that this is the case if, and only if, the Riemann curvature tensor is zero. Here the Riemann tensor is fully written out. However, the "Commentatio" probably played a minor role, if any, in the early history of GR; major developments were independent of this contribution.

### Christoffel: Algebraic Invariants

In his paper, Christoffel (1869), Elwin Bruno Christoffel formulated a more general version of Riemann's problem in the "Commentatio". Christoffel asks under which condition we can transform two quadratic differential forms into each other. In slightly modernized notation, his problem, the equivalence problem of homogeneous quadratic differential forms, concerns under which conditions the following equality holds:

$$g_{ik}dx_i dx_k = g'_{ik}dx'_i dx'_k \tag{5.17}$$

Why is this question relevant? In geometrical, slightly anachronistic terms, the question is whether the metric on the left is "the same" as the metric on the right. Riemann showed that the metric has $n \cdot \frac{n+1}{2}$ independent components, in total. If we change coordinates, this will only affect $n$ of these, while $n \cdot \frac{n-1}{2}$ are independent, i.e. they "belong to the metric". Christoffel asked under which conditions these $n \cdot \frac{n-1}{2}$ components of two metrics are connected by coordinate transformation, or if we can reach $g'_{ik}$ from $g_{ik}$ with a change of coordinates.[18]

In a nutshell, Christoffel's main result is that the equivalence problem depends only on the Riemann tensor and its derivatives. The problem is thus reduced to the equality of these quantities. However, in Christoffel's case, the path to the result is as important as the result itself. Along the way Christoffel introduces a number of notions, notations and techniques that will prove to be crucial in the further course of tensor calculus. He discovered the Riemann tensor, the eponymous Christoffel symbols, and also the expression that would later be interpreted as the covariant derivative.

[18] It has to be noted that, while Christoffel cites Riemann once, Riemann's influence on Christoffel is probably negligible. In particular, Christoffel derived the Riemann tensor independently of Riemann – he could not have known about Riemann's "Commentatio", which appeared in print only in Riemann's collected works; see e.g. Reich (1994) on this point.

Christoffel's importance for the further course of events cannot be overestimated. Jürgen Ehlers (1981), in a review of Christoffel's 1869 paper, and later work on the equivalence problem, assesses their role as follows:

> This review may show to what extent Christoffel's work on the equivalence problem has paved the way towards solving that problem and to characterize pseudo-Riemannian spaces intrinsically: He developed an analytic apparatus which, apart from notational simplicity, anticipated all essential ingredients of the tensor calculus of Ricci and Levi-Cevita [sic]; he introduced the first and most important non-tensorial, geometric object (of second order) – his set of three-index symbols; he recognized the basic role of systems of total differential equations and their integrability conditions of arbitrary order as a tool for differential geometry and applied them particularly when these systems are not completely integrable. (Ehlers, 1981, p. 533)

Two things about Christoffel's role, in the history of GR, are particularly noteworthy. First, many of the central notions were introduced by Christoffel, but his discussion often remained at a technical level; he did not make an effort to interpret his innovations and place them in a broader mathematical context. This carries over to the second important point: the total lack of geometrical notions in Christoffel's paper. The equivalence problem is treated as a problem in its own right, and while the relevance of the problem, to ideas by Riemann (and Gauss), is mentioned very briefly; Christoffel does not elaborate on the consequences of his result for differential geometry.

### Ricci & Levi-Cività: Absolute Differential Calculus

The culmination point of the previous mathematical developments, and probably the entry point for Einstein and Grossmann into the mathematical literature, is Ricci and Levi-Civita (1901), the now-famous survey paper on the "Absolute Differential Calculus and its applications" ("Méthodes de calcul différentiel absolu et leurs applications"), ADC for short. This paper introduces what we now would call tensor calculus and its applications – without, however, using the word "tensor".

In the first chapter, Ricci & Levi-Cività introduce the "algorithm" of the ADC. The goal is to develop a calculus of mathematical objects that are independent of the choice of variables – the form of these objects, tensors, is invariant under change of variables. In the introduction, Ricci & Levi-Cività point out that the line element, and the metric, are the fundamental object in this context, as they serve to give an intrinsic description of $n$-dimensional manifolds.

The chapter introduces the central notions of tensor calculus, including the distinction between co- and contravariant "systems" (tensors), tensor algebra and tensor calculus. Ricci & Levi-Civitá introduce covariant differentiation, a tensorial notion of derivative that is independent of coordinate systems; if we apply covariant differentiation to a tensor, a new tensor of higher rank results. They also discuss the Riemann tensor, but only very briefly; more on this central concept follows in their chapter three, on analytic applications. This chapter takes many of the techniques introduced by Christoffel, and turns them into a proper calculus.

Chapters 2-6 are supposed to demonstrate the variability and strength of the calculus, showing how the algorithm can help to solve problems in intrinsic geometry, analysis, geometry, mechanics, and physics.

In the introduction, Ricci & Levi-Civitá discuss the methodological goal of the paper, as well as the mathematical traditions in which they stand. They discern two major influences. On the one hand, they emphasize that the calculus is "entirely" due to Christoffel. On the other hand, the motivation for developing the calculus is attributed to the "genius of Gauss and Riemann". The ADC paper is truly a culmination of previous mathematical work, in the tradition of Christoffel, an algebraic, invariant-theoretic approach – as well as differential geometry, championed by Gauss and Riemann.

### 5.2.4 Einstein's Heuristic and Research Strategies

We have now seen the questions that motivated the search for a generalized theory of relativity, and we have a rough overview of the relevant physics and mathematics that served as a starting point for the search. On top of this, however, Einstein needed a heuristic that gave his search for the field equations of GR further constraints and guidance. Renn and Sauer (2007) identify four heuristic principles, "relatively stable structures", as they call them, that could have played this role. These principles, grounded in the physical and mathematical knowledge of the time, indicated how the quest for GR might proceed at the same time. Renn and Sauer conjecture that Einstein had two different research strategies, based on these four principles. We will first introduce the four principles, and then the two strategies.

#### Equivalence Principle

This is probably the most famous of Einstein's principles. It postulates an equivalence between a homogeneous static gravitational field, i.e. a field where gravitation is evenly distributed and does not change over time, and a uniformly and linearly accelerated frame of reference without a gravitational field: the motion of a massive body in these two situations is the same, assuming that we use an inertial frame in the first situation. This equivalence is incorporated in Einstein's famous "elevator model". A second important model, in the same spirit, is the "bucket model", which equates a uniformly rotating frame of reference with a stationary gravitational field. These principles are properly heuristic in that, literally speaking, their validity is very limited; for example, the first equivalence stated above only holds locally.

Both these models can be interpreted as instantiations of the more general thought that there is an equivalence between gravitational and inertial mass. Both models were at the center of Einstein's early investigations of special cases of the equivalence principle; in the second phase of the search for GR, the principle served both as a control mechanism and as a construction principle, depending on the research strategy he pursued.

### Generalized Relativity Principle

This principle is closely related to the equivalence principle. It suggests generalizing the idea from SR that there are no privileged inertial frame to non-inertial frames. In this sense, the principle has to do with the question as to which properties of space-time should be accepted a priori. However, in application, the principle appears to have taken a distinctively mathematical flavor. In Einstein's mind, the requirement to generalize SR was closely related to finding an appropriate mathematical description of gravitation that was independent of the chosen coordinate system. After this mathematical theory, the ADC, had been found, the role of the generalized relativity principle revolved around the question of the right mathematical formulation of the field equations – generally covariant or not? – and also the question as to how general covariance had to be interpreted.[19]

### Conservation Principle

The conservation principle has its roots in both classical mechanics and SR. In classical mechanics, quantities such as energy and momentum are conserved. SR brought about conceptual changes, showing that these two quantities could be transformed into each other. This led to a new, integrated conservation law, and a reformulation of the integrated quantities in the stress-energy-momentum tensor. This tensor, described above, enters into the field equation as the correct description of the source, i.e. the distribution of mass-energy-momentum that gives rise to the gravitational potential. The conservation principle expressed the expectation that a form of the conservation laws of classical mechanics, and SR, would be recovered in GR. It was not clear from the beginning whether an acceptable form of energy-momentum conservation had to be postulated as an additional requirement, or whether, as it turned out in the end, it would follow from the field equations. However, it was clear that a form of the divergence equation

[19] Generally covariant equations are equations that do not change form under a general class of coordinate transformation, e.g. differentiable coordinate functions.

was a necessary condition for acceptable field equations, be they generally covariant or not.

### Correspondence Principle

The correspondence principle is the requirement that the new, relativistic theory of gravitation should incorporate, in some form or other, the content of classical gravitational theory, and other well-established theories. More specifically, the expectation was that the new field equation of GR would yield, in some suitable limit (low velocities and weak fields), the classical field equation of gravity: the Poisson equation. It was expected that the classical limit would be reached via an intermediate, special-relativistic limit; we will discuss Einstein's expectations of how this works below.

The Poisson equation played a double role in the genesis of GR: On the one hand, it served as a starting point for the "modeling exercise" of GR. On the other hand, it could be used to check the correctness of candidate field equations.

The correspondence principle is the strongest of the heuristic principles in that it could not really be avoided: the new theory would not be acceptable if it were not able to explain, or incorporate, classical gravitational theory, which is, after all, accurate and well confirmed, with few exceptions.

### Two Research Strategies

The four heuristic principles do not uniquely determine the outcome of the search for GR; they are not even satisfiable simultaneously. This is not problematic *per se*, as they should not be read as axiomatic principles that Einstein never questioned; instead they changed content as Einstein's search progressed, and were able to take on different roles at different stages. Renn and Sauer (2007, sec. 3.4) discern two different research strategies that result from using the four heuristic principles in changing roles, either as "constructing principles", i.e. starting points of the investigation, or as "validity criteria", i.e. for checking the adequacy of candidate formulations.

First, the *physical strategy* is to take the existing physical theories as the starting point, and to find a suitable modification thereof. The physical strategy takes the correspondence principle as the starting point; candidate field equations are modeled on the basis of the Poisson equation. Candidates are then first checked against the physically motivated conservation principle. Only after passing this test would there be an attempt to determine the covariance group of the field equations, i.e. the extent to which the generalized relativity principle was satisfied, paying special attention to both the "elevator model" and the "bucket model" of the equivalence principle.

The *mathematical strategy*, on the other hand, takes the distinctively more audacious approach of starting from the principle of generalized relativity: suitable candidates for the field equations are taken from the mathematical knowledge about generally covariant differential operators. Only then are the physically motivated principles, first and foremost the correspondence principle, used to check the adequacy of candidates. This strategy is dubbed mathematical because the mathematics used in the construction had not been applied in this physical context before, and the exploration of its physical interpretation caused much more difficulties than for the physical strategy with its well-known mathematics.

These two strategies were not applied simultaneously, but in sequence. In the first stage of the genesis of GR, Einstein pursued the physical strategy by implementing the "elevator model" and the "bucket model" mathematically. Then, when these explorations suggested an extension of the mathematical framework, he switched to the mathematical strategy and checked candidate operators for the field equations. After he had (erroneously) convinced himself that this approach was not paying off, he returned to the physical strategy. This is, of course, a very rough account, and we will paint a more detailed picture below.

Renn and Sauer emphasize that the research strategies played an important role in organizing the available physical and mathematical knowledge. All the principles were connected with specific and varying physical and mathematical content. For example, the correspondence principle had to be implemented using the Poisson equation, as well as differential operators, as results of the weak-field special-relativistic limit of candidate generally-relativistic field equations, while an entirely new kind of mathematics had to be explored to implement the principle of generalized relativity. The sheer mass of available knowledge made it almost impossible to reconcile all the different parts at once. The two research strategies alleviated this problem, to a certain extent, by stressing different (mathematical and physical) parts of the available body of knowledge, thus making the task more tractable.

## 5.3 From the Beginning to the Entwurf Stage (1907 – 1912)

In Stachel (2007), John Stachel structures the genesis of GR as a three-act drama. We will follow Stachel's example and sketch the first two acts of the drama of GR, in section 5.3.1, drawing on our exposition of the main protagonists (predecessor theories), and the motives that may have guided the action of the play (heuristic principles and research strategies). The first act is Einstein's formulation of the equivalence principle in 1907. The second act tells the story of how the general, non-Euclidean metric became one of the central objects of gravitational theory, representing the gravitational, potential before 1912.

In section 5.3.2, I will give a quick overview of the first scene of act three, culminating in the Entwurf theory. This is the part of the drama we understand best, as its genesis is documented in the Zurich notebook. The end of the drama is the story of how Einstein found the correct field equations in 1915. We will turn to this part in sections 5.4 – the explanation for the three-year delay – and 5.5, the resolution and discovery of the final field equations.

### 5.3.1 The First Two Acts (1907 – 1912)

#### Act I

The main event of the first act is the introduction of the Equivalence Principle. The Equivalence Principle is, roughly, the extension of the Relativity Principle from SR to accelerated frames of reference; see section 5.2.4 above. Stachel notes that the implementation of the physical insight, that gravitational and inertial mass are equivalent, was delayed for several years because the appropriate mathematical concept for capturing this equivalence – an affine connection on a four-dimensional manifold – was not yet available.

#### Act II, Scene 1

In the first scene of Act II, Einstein realized around 1907 that even in the simple case of a uniformly accelerating reference frame, the "elevator model" of the Equivalence Principle, coordinates lose their direct meaning: while spatial measurement does not pose any problems in this model, the time coordinate no longer has a direct interpretation, as mirrored in the distinction between universal and local time.

The reason for this is that the rate of clocks is affected by gravitational fields (gravitational frequency shift).[20] In a uniform gravitational field, one case of a static field, it is possible to find a "universal time", which expresses simultaneity of distant events. However, in order to get such a universal time in a spatially varying gravitational potential, one has to adjust the rate of clocks according to potential differences. If we set a clock at an arbitrary point $A$ to be the standard clock that ticks at standard rate, the rate of a clock at a point $B$ will have to be multiplied with a (constant) factor $e^{-\phi_B/c^2}$, where $\phi_B$ is the potential at $B$ (we have set the potential at $A$ to zero), in order to bring the clocks at $A$ and $B$ to the same (universal) frequency.

Probably as early as 1909, Einstein also noted that spatial coordinates might lose their direct meaning as well. The implementation of the "bucket model", involving a uniformly rotating disk and frame of reference, questioned not only the direct interpretation of coordinates, but more generally the appropriateness of Euclidean geometry. Additionally, the behavior of a light ray in a gravitational field, corresponding to a uniformly rotating

[20] The following draws on Rindler (2006), in particular sections 1.16, 9.1-9.4.

disk, suggested that the field exerted a velocity dependent force on the light, meaning that the theory could not be scalar.

Here is a short explanation of the rotating disk model. The central role of this model in the genesis of GR was first pointed out in Stachel (1980). Stachel writes that the rotating disk could be the "missing link" between "flat" generalizations of SR and the Entwurf theory with a general metric. However, he notes that there is no clear historical evidence for the central role of the rotating disk; we can only infer it from later accounts by Einstein.

The most detailed account of the model can be found in a letter from Einstein to Petzold in 1919. In an earlier letter to Einstein, Petzold had claimed that if we measure lengths of a disk in uniform rotation, the circumference $C$ is length-contracted, while the diameter $D$, which is orthogonal to the direction of motion, is not. Therefore, $C < D \cdot \pi$, in contradiction to Euclidean geometry.

In his reply, Einstein disagrees and proposes the following analysis. We can measure diameter and circumference of the uniformly rotating disk in two frames of reference. Frame $K_0$ is at rest, and the disk rotates relative to it. If we measure diameter and circumference, with rods at rest in $K_0$, then their proportion is $\pi$. Now we choose a frame $K$ that is co-rotating with the disk, and we use rods at rest in $K$. Seen from $K_0$, the rods measuring the diameter are not length-contracted, wherease the rods measuring the circumference are. However, according to the measurement with these rods, the circumference is *longer* than measured in $K_0$: $C > D \cdot \pi$. In any case, Euclidean geometry can no longer hold.

Stachel argues that this argument suggests that, if we can treat the rotating disk with the means of SR, if the frames of reference can be brought in correspondence with gravitational fields, and if gravitational fields do not affect length measurement, rigid bodies cannot be adequately captured by Euclidean geometry.

Summing up, after scene 1, Einstein was probably aware of the fact that in two separate models, temporal and spatial coordinates lose their direct interpretation, and he might have realized that Euclidean geometry will not do, at least for spatial coordinates.

### Act II, Scene 2

Act II, Scene 2 is centered around the notion of gravitational potential. Stachel first notes that Max von Laue may have helped Einstein realize the importance of the gravitational potential. In a letter to Einstein at the end of 1911, von Laue pointed out that, in generalized relativity, the potential acquires direct physical meaning: it is measurable because of its influence on the speed of light. We have seen in the last section how the potential $\phi$ influences the rate of clocks in a gravitational field. The idea was to interpret the potential as the speed of light; the latter would depend on spatial position:

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2(x, y, z)dt^2 \tag{5.18}$$

Here $c(x, y, z)$ acquires the role of the static gravitational potential, $\phi(x, y, z)$. In two 1912 papers, Einstein started to explore this line of thought, in the case of static gravitational fields. The crucial insights may have been that we can interpret $c$ as a (variable) component of the metric, and von Laue's remarks on the correct representation of the potential, in the case of more general fields. If this is combined with the insight that the spatial coordinates might lose their direct meaning, and that they could be variable in the general case, generalizing the Minkowski metric, to a metric with variable coefficients, does not seem an outlandish option.

At the end of Scene 2, a further important tool was introduced. In a supplement of the second 1912 paper, Einstein derived the equation of motion from a Lagrangian formulation (see section 5.2.2 above), writing

$$\delta\left\{\int \sqrt{c^2dt^2 - dx^2 - dy^2 - dz^2}\right\} = 0 \tag{5.19}$$

The role of the Lagrangian formulation of the equations of motion is, first, due to the insight that the formulation is independent of any specific choice of coordinates. Second, it might have suggested a parallel to differential geometry to Einstein, as we will see in the next scene.

### Act II, Scene 3

Scene 3 describes the transition from the previous two scenes to the Entwurf theory, which is based on the ADC and uses the generally covariant metric. Stachel identifies two different kinds of influence that may have led to this formulation. On the one hand, there is, from the mathematical side, differential geometry, and in particular Gaussian surface theory. On the other hand, the physics literature also discussed the line element, especially in the context of rigid motion and bodies.

Einstein was familiar with differential geometry through lectures by Carl Friedrich Geiser. This part of the literature provided ideas about the use of arbitrary coordinate functions, the invariant line element $ds$, and geodesics, although only for two-dimensional surfaces. We saw above that Einstein used variational techniques to derive the equation of motion in the supplement of his second paper on static gravitational fields. In the Entwurf theory, Einstein switched to writing $ds$ for the square root of the line element, as in equation (5.10), thereby opening the path to a new geometric interpretation of the equation of motion: the variation can now be interpreted as giving rise to the motion of a particle on a geodesic in a non-Euclidean space-time, in close analogy to geodesics on two-dimensional surfaces in Gaussian surface theory (see section 5.2.3 above).

The root of this interpretation is, first, the line element in Gaussian surface theory, according to which geodesic lines in a Gaussian surface are given by equation (5.15). Second, the interpretation of the Minkowski metric as line element had entered the physical discussion before, in the context of rigid bodies in SR. Stachel writes that the similarity of the geodesic equation from differential geometry to the equation of motion in a variational formulation "could have suggested the analogy between Gauss' theory of surfaces and Einstein's theory of the static gravitational field" (Ibid., p. 104). Stachel sums up scene 3 as follows:

> [O]n the basis of the mathematical and physical resources at his command, at some point in mid-1912, after generalizing the single gravitational potential $c$ to the array of ten gravitational potentials $g_{ik}$, Einstein realized that they formed the coefficients of a quadratic form $\sum g_{ik}dx_i dx_k$ which could be regarded as the square of the invariant line element ($ds^2 = \sum g_{ik}dx_i dx_k$) of a four-dimensional spacetime manifold; and that the interval $ds$ represents a physically measurable quantity—the proper time if the interval between two events were time-like, the proper length if it were space-like (of course it would vanish for null intervals) (Ibid., p. 106)

With all of this at hand, Einstein turned to Grossmann, who introduced him to the generalizations of Gauss by Riemann, Christoffel, Ricci & Levi-Cività. Stachel notes that, unfortunately, despite the parallels with Gaussian surface geometry, Grossmann did not take a very geometrical route in his 1913 discussion of tensor calculus.

### 5.3.2 Act III, Scene 1: The Entwurf Theory

This section contains a short description of the state of the drama at the beginning of Act III.[21]Einstein had adopted the "Lorentz model" for GR. His main goal was to find field equations of gravitation, which describe how the gravitational field is generated locally, depending on the distribution of mass-energy and momentum. In a second step, he wanted to derive an equation of motion, from the field equations.

#### The Field Equations

Renn and Sauer represent the model of a field equation symbolically as follows:

$$OP(POT) = SOURCE \tag{5.20}$$

[21]A detailed account of the story can be found in Renn and Sauer (2007).

This is a general model of a field equation and has different instantiations depending on context. One example is the Poisson equation (5.3) from classical mechanics. The model has three components. First, the SOURCE slot has to be filled with a term for the source of the gravitational field. Second, the POT slot has to be filled with an expression for the potential. Third, the OP slot has to be filled with a differential operator acting on the POT term. In classical mechanics, SOURCE is matter density $\rho$, POT is the classical potential $\phi$, and OP is the Laplacian $\Delta$.

In the beginning of Act III, Einstein had already settled for generally-relativistic candidates of both POT and SOURCE. The POT slot was instantiated by the generally covariant metric tensor $g_{\mu\nu}$. We have seen how the Minkowski metric was generalized in the first two acts of the drama; its identification with the potential was described in section 5.2.2.

The SOURCE slot was instantiated by the Energy-Momentum tensor; see section 5.2.2 above.[22] There were physical reasons, coming from classical mechanics and SR, for choosing the EM tensor as the appropriate choice for SOURCE in GR. In the context of the field equation, a further, mathematical reason is the form of equation 5.20: the metric tensor is a two-index tensor, which suggests that the right-hand side of the equation should be instantiated with a two-index tensor as well. This further stabilized the choice of both SOURCE and POT.

What remained, then, was the question of how to instantiate the differential operator for the OP slot. As we will see, this final part of the puzzle posed the greatest problems. Other than in the case of SOURCE and POT, it was not clear what would be a suitable candidate for OP, other than the requirement that it should generalize the Laplacian operator from the Poisson equation. It is presumably at this point that Einstein turned to his "mathematician friend" Marcel Grossmann for help.

### Grossmann's Role in the Genesis of GR

Marcel Grossmann and Albert Einstein both studied at the ETH Zürich; Einstein took physics, while Grossmann focused on mathematics.[23] Their coursework overlapped, and classes were small, so they became friends. Years later, Einstein remembered that he used Grossmann's carefully transcribed lecture notes to learn for the exams – including the notes of Geiser's lecture on infinitesimal geometry ("Infinitesimalgeometrie"). This lecture could be responsible for the momentous shift towards the application of geometry with variable curvature in physics; see section 5.2.3 above.

Grossmann was appointed professor of mathematics at the ETH in 1907. By 1911, he had become a heavyweight in the ETH, and in 1911, he worked

[22] In the so-called source-free case, the SOURCE term is identically zero, i.e. the EM-tensor vanishes.

[23] This subsection is based on Sauer (2013).

towards an engagement of Einstein, taking over the now-vacant professorship previously held by Hermann Minkowski. In January 1912, Einstein was appointed professor of theoretical physics, effective fall of 1912. Einstein moved from Prague to Zürich in the summer of 1912, and started a collaboration with Grossmann on GR soon afterwards. The collaboration lasted until Einstein left for Berlin in 1914.

The collaboration between Einstein and Grossmann is not only of historical interest: it is also fruitful for the better understanding of philosophical issues surrounding the application of mathematics in empirical science. We can interpret the collaboration of Einstein, the physicist, and Grossmann, the mathematician, as an exemplification, and personification, of the interaction of mathematics and physics, in one of the most important episodes in the history of science. The hope is that, if we understand Grossmann's contribution to the genesis of GR, we also gain a clearer picture of the role of mathematics in application.

The question we have to answer is what Grossmann's contribution to the genesis of GR was. Unfortunately, it is not clear what the exact state of research was, or what Einstein already knew, when he approached Grossmann. In later recollections, Einstein gives fairly detailed accounts; however, they are incomplete, and not very reliable. In 1955, Einstein described the state of research, and the question he asked Grossmann, as follows (quoted after Sauer (2013, p. 8)):

> The problem of gravitation was thus reduced to a purely mathematical one. Do differential equations exist for the $g_{ik}$, which are invariant under nonlinear coordinate transformations? Differential equations of this kind and only of this kind were to be considered as field equations of the gravitational field. The law of motion of material points was then given by the equation of the geodesic line. With this problem in mind I visited my old friend Grossmann who in the meantime had become professor of mathematics at the Swiss polytechnic. He at once caught fire, although as a mathematician he had a somewhat skeptical stance towards physics.

If this account is correct, then Einstein's challenge was to find candidate differential operators instantiating OP in the field equation. However, it is not clear whether, at this point, Einstein already had a "tensorial" approach in mind, i.e. whether he wanted the differential operator to be generally covariant. Maybe this was Grossmann's suggestion.

Besides these recollections, there are several contemporary documents that give us insight into the collaboration; two of them stand out. One is the so-called Zurich notebook: Einstein's notebook containing research notes on GR (among other things), written between the summer of 1912 and spring

1913. The second is the Entwurf theory, Einstein and Grossmann (1995), the first published account of GR based on the ADC.

Both documents are telling in their own way. The Zurich notebook gives us direct insight into the process of discovery, the laboratory, of Einstein's work on GR. Here Einstein examined various approaches to the mathematical formulation of GR, and surprisingly, he wrote down a (linearized) version of the final, correct field equations years before their "true" discovery. The notebook is entirely in Einstein's hand. However, Grossmann's name appears at critical junctures. The relevant pages suggest that it was Grossmann who familiarized Einstein with the new mathematics. His name appears in conjunction with the Riemann tensor, and another differential expression.[24] We will not analyze the role of the Zurich notebook in the context of Grossmann's contribution here; we hope to make good for this lacuna at a later point.

The Entwurf paper is the document we will focus on in chapter 6. It has two parts. The first, physical part is authored by Einstein; the second, mathematical part is written by Grossmann.

Einstein's part presents much of GR as it is known today. He introduces the invariant line element as a measure of distance between infinitesimally close space-time points using the metric $g_{\mu\nu}$ (the invariance holds for arbitrary variable substitutions); he derives the energy-momentum balance equation, using a variational principle and the energy-momentum tensor for dust; and, most importantly, states and discusses the question of how to find a field equation, as a generalization of the classical Poisson equation.

At this point, however, the Entwurf theory deviates from the modern route. Einstein writes that it is impossible to find a generally covariant differential operator that enters into a potential field equation and reduces to the Poisson equation in a suitable manner; see Einstein and Grossmann (1995, p. 312). He refers the reader to the second, mathematical part for the argument. Instead of taking the route suggested by the mathematical theories now at his disposal – what Renn and Sauer call the mathematical strategy; see section 5.2.4 above – Einstein chose to construct a field equation that is not generally covariant; he followed the physical strategy and derived the Entwurf equations using the conservation principle.

Einstein failed to see that it is in fact possible to construct a generally covariant field equation, which delayed the formulation of GR in its final form for three years. The reason for this failure cannot be understood from the Entwurf paper alone; it is necessary to consult the Zurich notebook. We will return to this part of the story below.

In the mathematical part of the Entwurf, Grossmann lays the mathematical foundations for the first tensorial formulation of a theory of gravitation.

[24]The notebook is now a well-understood part of the history of GR. A facsimile and reconstruction is given in the "Genesis" volumes.

He begins with a programmatic introduction, in which the most important mathematical sources he used, Christoffel (1869) and Ricci and Levi-Civita (1901), are cited. First Grossmann gives an exposition of tensor algebra and tensor calculus. The most important part of the paper is part four, the "service part" for the physical theory. Grossmann proves that the energy-momentum balance equation is generally covariant, discusses the generally covariant approach to the field equations, and finally provides the mathematical derivation of the Entwurf field equations.

In chapter 6, our goal will be to understand the transition from the purely mathematical theories, introduced in section 5.2.3, to Grossmann's part of the Entwurf. What is the origin of the mathematics introduced in Grossmann's part of the Entwurf? And: how did he transform the mathematics to adapt it to the physics? We will have a detailed look at Grossmann's part, trace the origins of his ideas, and discern his own contributions. This historical groundwork will help us understand the systematic question of how (pure) mathematics is transformed for applicability, in chapter 8.

## 5.4 Act III, Scene 2: Progress in a Loop (1912 – 1913)

Act III, Scene 2 of the drama is the last twist before the great resolution; it is centered around the question as to why it took Einstein three more years to complete the theory and settle on the Einstein field equations. We have already touched on some of the major questions: Why did Einstein and Grossmann abandon the mathematical strategy in the Entwurf? Why did they believe that the Ricci tensor does not yield the right classical limit? This part of the story is discussed in section 5.4.1.

However, the problem with the Ricci tensor is only the tip of the iceberg. The so-called "November tensor", a further candidate differential operator, poses an even greater puzzle. The November tensor did not face the same difficulties as the Ricci tensor, but it was nevertheless rejected around 1913, only to be briefly revived in November 1915. There is not yet a definitive account of what the problem with this tensor may have been, but the discussion reveals further conceptual issues regarding the interplay between mathematics and physics. We will discuss this part of the story in section 5.4.2.

It is necessary to draw on recent results, based on Einstein's Zurich notebook, to reconstruct Einstein's odyssey. Such a reconstruction is available in the form of the "Genesis of General Relativity" volumes, in particular Janssen et al. (2007a,b). The brief account given here is mainly based on Renn and Sauer (2007); Norton (2007); Janssen and Renn (2007), as well as the accessible discussion in Norton (2005).

### 5.4.1 Rejecting Ricci

The first question we want to answer is: what exactly went wrong with the classical limit of the Ricci tensor? Before the analysis of the Zurich notebook, it was commonly held that Einstein and Grossmann were simply not aware of a modern, standard feature applied to obtain the Newtonian limit, so-called coordinate conditions; Pais (1982, p. 222) is an example.

Coordinate conditions are conditions imposed on the coordinate systems in order to recover the Newtonian limit. Coordinate conditions, such as the harmonic coordinate conditions we will encounter below, are necessary to recover the Newtonian limit from the generally covariant field equations; this results in a restriction of covariance to Galilean coordinate transformations. Galilean coordinate transformations are closed under coordinate systems that are in constant relative motion. They are only used in the context of the classical limit, and do not restrict the covariance group of the generalized field equations.

The Zurich notebook shows that Einstein was aware of the *mathematical* possibility of imposing coordinate conditions. For example, he invoked harmonic coordinates in order to recover the Laplacian operator from the Ricci tensor. An analysis of the Zurich notebook suggests that two related misconceptions prevented Einstein from recognizing the Ricci tensor as a viable candidate differential operator for the field equations.

The first misconception was Einstein's expectation that static fields are spatially flat. In a *static* gravitational field, the metric tensor can be given the following form in certain coordinates:

$$\begin{bmatrix} g_{11} & g_{12} & g_{13} & 0 \\ g_{21} & g_{22} & g_{23} & 0 \\ g_{31} & g_{32} & g_{33} & 0 \\ 0 & 0 & 0 & g_{44} \end{bmatrix} \tag{5.21}$$

A space-time of this form has the property that time and space components can be considered in isolation; there is no interaction between time and space coordinates. We can recover the usual notion of three-dimensional space, with its geometrical properties, by running through the time coordinate.

A metric is *spatially flat* if it can be transformed into the following form:

$$\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & g_{44} \end{bmatrix} \tag{5.22}$$

The only non-constant component of this kind of metric is $g_{44}$. In other words, Euclidean geometry is valid for the spatial components.

We have seen above that Einstein had already explored the form of spacetimes that implement the principle of equivalence in simple cases, in particular coordinates in uniform linear acceleration, which are equivalent to a static, homogeneous gravitational field. In such cases, the line element takes the form 5.18, and all but the time component of the metric can be transformed away. Einstein failed to see that the homogeneous case is only a special case of static fields in which spatial curvature vanishes.

The second misconception had to do with Einstein's expectation as to how the Poisson equation would be recovered. To obtain the Poisson equation in the classical limit, one considers the case of *weak fields*, in which there are only small deviations from the Minkowski metric $\eta_{\mu\nu}$; they can be written as follows:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{5.23}$$

where $h_{\mu\nu} << \eta_{\mu\nu}$, and derivatives of $h$ are small. Einstein assumed that the differential operator, entered into the field equations, would reduce to

$$\Gamma_{\mu\nu} = \sum_{\alpha\beta} \frac{\partial}{\partial x_\alpha}\big(\gamma_{\alpha\beta}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\big) + \big(\text{second order terms}\big) \tag{5.24}$$

This is a simple, generally covariant generalization of the Laplacian operator. $\gamma_{\mu\nu}$ is the contravariant form of the metric. Note that we assume the determinant of the metric to be 1. The idea of the second summand is that, if the deviations from the Minkowski metric are only small, then products of these deviations, terms of second order or higher, will be tiny, and can be neglected. If this operator is inserted into the field equations, and the special case of a spatially flat metric is considered, along with pressureless, motionless dust as the energy-momentum tensor[25], then the field equations essentially reduce to the Poisson equation.

If equation 5.24 were indeed the right intermediate step, between the general field equations and the classical limit, and all but first order terms were retained, then, in the weak field limit, we would get

$$\Box g_{\mu\nu} = \kappa\Theta_{\mu\nu} \tag{5.25}$$

where $\Box$ is the d'Alembertian, the special-relativistic generalization of the Laplacian. However, it turns out that in the final theory, the weak field limit contains a trace term[26]:

$$\Box g_{\mu\nu} = \kappa(\Theta_{\mu\nu} - 1/2 g_{\mu\nu}\Theta) \tag{5.26}$$

Einstein's expectations for the weak, and the static, fields are related: If one solves equation 5.25 in the special case of a time-independent field, i.e.

[25] See section 5.2.2 above.

[26] The trace of a rank two tensor is the sum of its diagonal elements.

a field that does not change over time, and pressureless, motionless dust as the energy-momentum tensor, then one recovers a spatially flat metric. This inference is blocked if one takes the route via equation 5.26, i.e. the weak field equation with a trace term.

How did these two misconceptions lead to the rejection of the Ricci tensor? It is possible to reconstruct Einstein's path of reasoning on the basis of some pages of the Zurich notebook. If one accepts Einstein's premises, about weak and static fields, then the Ricci tensor has to be rejected. In order to reach the Newtonian limit, one has to introduce coordinate conditions, so-called harmonic coordinates:

$$\sum_{\kappa l} \gamma_{\kappa l} \begin{bmatrix} \kappa l \\ i \end{bmatrix} = 0 \tag{5.27}$$

This condition makes it possible to recover the Newtonian limit from equation 5.24, as the superfluous second derivatives in the left-hand term in this equation are set equal to zero. However, if one applies the harmonic coordinate conditions to a weak, static field, one does not recover the spatially flat metric 5.22, as Einstein expected – a different condition is necessary to achieve this goal.[27] For Einstein, this was sufficient for rejecting the Ricci tensor, and harmonic coordinate conditions no longer play a role in the Zurich notebook.

### 5.4.2 November Nullified

The Zurich notebook makes it possible to delve deeper into Einstein's conception of the interplay between mathematics and physics, in particular the evolution of his interpretation of coordinates, different covariance groups, and the interpretation of other aspects of the ADC. The goal of this subsection is to have a brief look at some of the issues that Einstein was struggling with between 1913 and 1916, which had to be resolved before he could formulate the final theory of GR.

One of the major puzzles in the genesis of GR is that Einstein rejected a second differential operator, the November tensor, in the Zurich notebook, while he thought for a short period in November 1915 that it would be an acceptable differential operator for the field equations. What was the reason for this change in perspective?

#### The November Tensor and the "Fateful Prejudice"

The November tensor is not a tensor of general covariance, it is only covariant under so-called *unimodular* transformations. These are transformations for which the matrix of the coordinate differentials, $\left[\frac{\partial x'_\beta}{\partial x_\alpha}\right]$, has determinant 1.

[27] In personal communication, Tilman Sauer speculated that harmonic coordinates and weak, static fields, together with spatial flatness, might be mathematically inconsistent.

The November tensor results from a decomposition of the Ricci tensor $T_{il}$ into two summands:

$$T_{il} = \left(\frac{\partial T_i}{\partial x_l} - \sum \begin{Bmatrix} il \\ \lambda \end{Bmatrix} T_\lambda\right) - \sum_{\kappa l} \left(\frac{\partial \begin{Bmatrix} il \\ \kappa \end{Bmatrix}}{\partial x_\kappa} - \begin{Bmatrix} i\kappa \\ \lambda \end{Bmatrix}\begin{Bmatrix} l\lambda \\ \kappa \end{Bmatrix}\right) \tag{5.28}$$

The first summand is a tensor of unimodular covariance, as it is the covariant derivative of a vector of unimodular covariance, the coordinate derivative of $log\sqrt{det(g_{\mu\nu})}$. The second summand, the November tensor, is therefore also unimodular, as it can be written as the difference between a generally covariant and a unimodular tensor. While unimodular covariance is a subgroup of general covariance, it still realizes the principle of equivalence in important cases, including rotations around spatial axes, and acceleration of the spatial origin.

The November tensor was attractive for two further reasons. First, if one applies the so-called "Hertz condition"

$$\sum_\kappa \frac{\partial \gamma_{\kappa\alpha}}{\partial x_\kappa} = 0 \tag{5.29}$$

to the November tensor and considers the weak field solution of the new object, the result agrees with Einstein's expected weak field equation 5.24, up to second order quantities. Second, the November tensor satisfies the requirement of energy conservation in the weak field form. Thus, if one takes the November tensor as a candidate differential operator, both problems with the Ricci tensor are resolved. Einstein played around with the November tensor in the Zurich notebook, and discarded it nevertheless. What were his reasons?

In retrospect, Einstein gave several accounts for why he abandoned the November tensor, attributing the decision to a "fateful prejudice". Firstly, he writes that he was unable to recover the Newtonian limit. Secondly, considerations of energy-momentum conservation led him to interpret the components of the metric, and not the Christoffel symbols, as the components of the gravitational field, as he thought later. His belief about the components of the gravitational field prompted him to expand the products of the Christoffel symbols in the November tensor, in order to get an expression in terms of derivatives of the metric. This calculation may have proven to be too difficult, or to be not simple enough.

One difference between the first exploration of the November tensor, and the situation in November 1915, was that Einstein had developed variational methods, which allowed him to establish energy-momentum conservation in an easier manner. Furthermore, if the "fateful prejudice" is abandoned, and the expression is written in terms of Christoffel symbols, and not directly in terms of the metric, the result is very simple.

However, Einstein's explanation as to why he dismissed the November tensor in the Zurich notebook seems odd. Is it reasonable to abandon this promising candidate simply because a calculation looks complicated? The remark that the November tensor did not yield the right classical limit also seems strange, in view of the fact that, in the notebook, Einstein showed how one can apply the Hertz condition 5.29, to obtain the weak field form of the operator 5.24 that he expected. Why did he abandon this approach nevertheless?

### Coordinate Conditions and Coordinate Restrictions

Einstein scholars have not reached an unanimous verdict on this issue; divergent explanations have been proposed. Here I will recount what appears to be the majority view. In the subsequent subsection, I will sketch an alternative explanation proposed by John Norton.

The majority view of what went wrong with the November tensor is based on two different notions of what it means to apply, say, the Hertz condition to a candidate differential operator. Firstly, these equations can be used as *coordinate conditions.* Coordinate conditions are a now-standard tool for obtaining the classical limit of the generally covariant field equations. Classical equations are only covariant under Galilean transformations. If one takes the limit of weak, static fields, the resulting equations are covariant under a bigger group of transformations. Therefore, it is necessary to impose further restrictions to recover Galilean covariance. This can be achieved with the help of coordinate conditions. Coordinate conditions are not a restriction on the generally covariant theory, but just a tool for recovering the classical limit.

Secondly, conditions, such as harmonic coordinates and the Hertz conditions, can be interpreted as *coordinate restrictions.* The idea behind coordinate restrictions is to limit the covariance group of some tensor $A$, in order to find a new tensor $A'$, of limited covariance, which is itself a candidate operator. One example of this is the November tensor itself. It is generated from the generally covariant Ricci tensor, and despite its covariance under unimodular transformations, it was considered to be a candidate operator for the field equations.

Coordinate restrictions are only acceptable if one is ready to give up on general covariance for the final theory and accept unimodular covariance instead. A possible justification for such a restriction could be that general covariance is too permissive, from the perspective of the equivalence principle, as some changes of coordinate systems do not correspond to genuine changes in the state of motion.

This raises the question as to how to interpret Einstein's calculations in the Zurich notebook. Did he conceive of the Hertz condition, or the harmonic coordinates, as coordinate conditions that only serve to recover the classical limit, or did he use them in the more substantial sense of coordinate restrictions? And if he used coordinate restrictions, was he aware of the possibility of using the same expressions as coordinate conditions?

There is evidence that Einstein interpreted the Hertz condition as a coordinate restriction, and not as a coordinate condition. A calculation in the Zurich notebook suggests that Einstein wanted to find the covariance group of the November tensor under the Hertz condition. This does not make sense if one is merely interested in the Newtonian limit of the November tensor. If, on the other hand, Einstein was searching for a differential operator conforming to his expectation for the weak field limit, then it was important to determine its covariance group. Einstein did not finish the calculation, but, if he did, he would have found that the resulting differential operator is not acceptable. It is not invariant under spatial rotations, a requirement that he checked on other occasions.

Further calculations in the notebook, using a different condition, the so-called "Theta requirement", also point towards a use of coordinate restrictions. In this case, Einstein checked whether the November tensor, together with the Theta requirement, is invariant under spatial rotations, and found that this is not the case. Again, this calculation would not make sense, had Einstein only been interested in recovering the Newtonian limit. It seems that he wanted to check covariance properties of the November tensor, combined with the Theta requirement.

These two calculations suggest that Einstein rejected the November tensor because he interpreted the harmonic coordinate condition, the Hertz condition, and the Theta requirement, not as coordinate conditions, but as coordinate restrictions. The requirements genuinely constrain the covariance group of the differential operator, and consequently the covariance group of the field equations. If a tensor, together with the coordinate restriction, did not conform to Einstein's expectations, it had to be abandoned.

Some of the requirements, such as the Theta requirement, seem to make sense only if they are interpreted as coordinate restrictions. In other cases, for example the harmonic coordinate condition, it is not possible to settle for a definite answer as to how it should be interpreted, based on the Zurich notebook. In the end, the question boils down to whether Einstein was aware of the fact that he could use coordinate conditions in the modern sense, or whether he consistently used coordinate restrictions. There is no agreement on this issue amongst Einstein scholars. In the next subsection, we have a brief look at a divergent interpretation of Einstein's reasons for rejecting the November tensor.

### Minority Report: Norton's Hole Argument

In Norton (2005, 2007), John Norton disagrees with the above account, which explains Einstein's rejection of the November tensor, based on the distinction between coordinate conditions and coordinate restrictions. He thinks that attributing a confusion between these two notions to Einstein is implausible, as Einstein never conceded committing this mistake later, and as it is a fundamental oversight at the heart of his expertise. Norton prefers to trace Einstein's dismissal of the November tensor back to a version of the famous Hole Argument. Here I will recapitulate this argument very briefly, and explain how, according to Norton, it explains Einstein's rejection of the November tensor.

Einstein originally proposed the Hole Argument[28], to establish that general covariance is not a desirable feature of field equations, by showing that generally covariant field equations are indeterministic. Here is a sketch of the argument.

Assume that the metric field is determined by the (source free) field equations $\Gamma_{\mu\nu} = 0$, and $g_{\mu\nu}$ is a metric solving these equations. As the field equations are generally covariant, we can express the metric in any coordinate system we like, say, as $g'_{\mu\nu}$ in primed coordinates. This means that we can let the metrics $g_{\mu\nu}$ and $g'_{\mu\nu}$ agree everywhere, except in some space-time region, the hole, where one deviates smoothly from the other. The solutions of $\Gamma_{\mu\nu} = 0$ will agree everywhere, except in the hole. This, however, implies that the metric outside the hole does not determine the metric inside the hole, which is an unacceptable form of indeterminism.

This argument against general covariance, the Hole Argument, is flawed. In order to understand why it is deficient, it is useful to introduce the distinction between active and passive transformations.

The $g_{\mu\nu}$ are the components of the metric *in one coordinate system* $x_\alpha$: there is a functional dependence of the components of the metric on the coordinates, $g_{\mu\nu}(x_\alpha)$.[29] In a *passive transformation*, we change the coordinate system, $x'_\beta(x_\alpha)$ (read: the primed coordinates are functions of the unprimed coordinates), and let the components of the metric co-vary, yielding new components expressed in a new coordinate system, $g'_{\mu\nu}(x'_\beta)$. Because of general covariance, this new expression is also a solution of the field equations.

The passive transformation $g'_{\mu\nu}(x'_\beta)$ is also a solution to the field equations, if we let the components of the metric be functions of the original coordinate system, $x_\alpha$. This yields an *active transformation* $g'_{\mu\nu}(x_\alpha)$: the primed components of the metric as functions of the old, unprimed coordinate system.

Why is the Hole Argument flawed? The problem is that the active transformation of the metric might suggest that, if we admit general covariance, it is possible to ascribe *different metrical properties*, $g_{\mu\nu}(x_\alpha)$ and $g'_{\mu\nu}(x_\alpha)$, to

[28] The following account is based on Norton (2005).

[29] The dependence can be seen explicitly in the expression of the Gaussian metric for two-dimensional surfaces in equation 5.14: $x, y, z$ are coordinate functions and change under coordinate transformations; the components of the metric, the functions $E, F, G$, change accordingly.

the *same* space-time point $x_\alpha$. This, however, is a mistake: If it is possible, as assumed in the argument, that we can transform $g_{\mu\nu}$ into $g'_{\mu\nu}$, using coordinate transformations, then they are physically the same. $g_{\mu\nu}$ and $g'_{\mu\nu}$ are mathematically different expressions, or different components, of the same class of metrics linked by coordinate transformations. It does not matter whether we use the primed or the unprimed coordinate system to express the components of the metric.

This also means that the coordinate functions $x_\alpha$ and $x'_\beta$ are not sufficient for picking out space-time events. It is only the combination of coordinates and metric that determines space-time properties of events. Two mathematically different descriptions of a space-time point are physically the same, if, and only if, we can transform one description of the point, the components of the metric in that point, into the other components, the other description.

How does the Hole Argument explain Einstein's rejection of the November tensor? Norton conjectures that Einstein may have ascribed an "independent reality" to some set of coordinates that mirror the structure of classical mechanics. One of the test cases of transformations, that he considered to be a necessary part of a restricted covariance group, were spatial rotations. Einstein may have considered the active transformation of the Minkowski metric under spatial rotations. However, this metric is not compatible with the Hertz condition; equation 5.29. Just as in the Hole Argument, this presupposes that it is possible to "remove" the Minkowski metric from a privileged set of coordinates, and introduce rotational coordinates afterwards. We now know that this does not make sense.

The advantage of Norton's account is that it is possible to explain Einstein's mistake on the basis of the Hole Argument, which he defended during the period in question; it is not necessary to ascribe a new, hitherto unknown, confusion to him.

## 5.5 Act III, Scene 3: Resolution (1913 –1916)

In this final section, I give a brief, somewhat cursory, account of Act 3, Scene 3, the events after the Zurich notebook and Entwurf phase described above, based on Renn and Sauer (2007, sec. 7).

After the publication of the Entwurf theory, Einstein was criticized because the Entwurf field equation was not generally covariant, and because it was unclear how the Entwurf theory was related to the generally covariant objects of the ADC. Einstein used the Hole Argument to defend the lack of general covariance of the Entwurf theory; he also adduced problems with energy-momentum conservation on the generally covariant approach, to argue against such a formulation.

In 1914, Einstein thought he had clarified the problem of how general covariance was related to the Entwurf equations. Working in a variational formulation, he claimed that, by postulating two requirements, one based on energy-momentum conservation, the other on the generalized relativity principle, the Entwurf Lagrangian could be shown to be unique. This argument was designed to support the Entwurf theory from the point of view of the mathematical strategy, while the previous formulation used the physical strategy, and in particular the conservation principle.

The realization that the Entwurf equations did not constitute an acceptable generalized theory of gravitation was a gradual process. The theory has three major problems. First, the Entwurf theory is unable to explain Mercury's anomalous perihelion precession; one of the few empirical phenomena that were not correctly predicted by classical mechanics. Second, the Entwurf equations did not contain the Minkowski metric in rotating coordinates as a solution; rotating coordinates were taken to correspond to a simple state of accelerated motion, rotation, according to the equivalence principle. Third, Einstein's "proof" that the Entwurf Larangian was unique under two reasonable assumptions proved to be erroneous. It is probably a combination of these problems that led Einstein to abandon the Entwurf theory. The last problem is the only one Einstein mentions in writing. He gave up the Entwurf theory some weeks after discovering the non-uniqueness.

In November 1915, he definitely returned to the mathematical strategy, and published a new theory, based on the November tensor. The fact that Einstein returned to candidate differential operators may seem puzzling at first. However, it was a reasonable course of action, if one takes into account that, first, certain properties of the candidates, especially energy-momentum conservation, had been left unexplored in the Zurich notebook. Also, Einstein had developed more powerful mathematical techniques to scrutinize the candidates. This, together with the fact that the November tensor had already been a promising candidate earlier, as it did not face problems in the weak, static field limit, is sufficient to explain Einstein's renewed interest.

Only a week after the publication of the November tensor, Einstein wrote an addendum, in which he returned to the Ricci tensor. He had to weight problems with the November tensor against problems with the Ricci tensor. The former implied an unmotivated restriction on general covariance, while he had found the latter to be problematic in the Zurich notebook, because it implied that the trace of the stress-energy tensor vanished, and thus contradicted his expectations for a theory of matter.

A week after this addendum, Einstein successfully explained Mercury's perihelion precession using the field equations with the Ricci tensor. This was, of course, a very strong empirical confirmation of the generally covariant approach. Einstein was able to carry out the calculation very quickly because he had already tried to explain the perihelion shift with the Entwurf theory; this attempt had been a quantitative failure. Luckily, the explanation of Mercury's perihelion does not depend on the still-missing trace term of the final field equations.

Now only a last modification was necessary to get from the field equations, based on the Ricci tensor, to the final, full field equations. Einstein realized that he could solve a problem with the energy-momentum balance, by adding a trace term of the energy-momentum tensor on the right-hand side of the field equation. This made it possible to abandon an additional, artificial requirement on the determinant of the metric. The resulting equations were a form of the Einstein field equations.

Of course, there were still many open questions; for example, a verification of energy-momentum conservation in the new theory. However, we will let the curtain fall on the drama of the genesis of GR at this point.

# Chapter 6

# Grossmann's Sources

## 6.1 Introduction: Motivations, Questions, Methods

The topic of this chapter is the historical question as to what Marcel Grossmann contributed to the genesis of GR.[1] We will examine the so-called "Entwurf" paper, an important joint publication of Einstein and Grossmann, containing the first tensorial formulation of GR. In particular, we will analyze the second, mathematical part of the Entwurf, and we will discuss the origin of the mathematical theories used in this part, as well as Grossmann's own, novel contributions.

The historical issues we discuss in this chapter are relevant for several philosophical issues. Our main systematic interest is in the general problem of how (pure) mathematics is applied to an empirical problem. The application of tensor calculus in GR is a prime historical case study for this problem. The very beginning of the tensorial formulation of GR is especially suitable, because the Entwurf theory mirrors the division of labour between mathematics and physics, in that the two authors, the mathematician Grossmann and the physicist Einstein, wrote their separate parts of the Entwurf.

The Entwurf theory constitutes the earliest meeting point the historical predecessor of tensor calculus, the "Absolute Differential Calculus" (ADC), and Einstein's generalized theory of gravitation. Previously, the ADC had been developed independently of application to gravitational theory. By comparing Grossmann's part with the mathematical theories he used, we can gain a better understanding of what is involved in the first steps of assimilating a mathematical theory to a physical question.

We will not explore these systematic issues in the present chapter, but will rather limit ourselves to the historical dimension of Grossmann's role in the early genesis of GR. We will put our historical insights to work in

[1] This chapter is based on joint work with Tilman Sauer. All translations are ours, unless stated otherwise.

chapter 8, in which we will confront the philosophical theory introduced in chapter 7 with the historical case discussed here.

There are several limitations to the historical width and depth of the present study. First, our focus is on the origin and transformation of those mathematical theories that had not been applied to gravitational theory prior to the Entwurf. We will not discuss the evolution of mathematics that evolved simultaneously to physics, especially the mathematical innovations in SR due to Minkowski, Sommerfeld, Laue, and others. These authors are also important for the Entwurf, but the influence of their tradition is much better understood than the purely mathematical tradition; see Norton (1992) for the distinction of the two traditions.

Secondly, we will neglect several documents that are relevant to Grossmann's contribution to GR. Most importantly, we will not examine Grossmann's role in the Zurich notebook, which documents the genesis of the Entwurf theory. We will also neglect other contemporary documents and later recollections. We hope to examine the Zurich notebook, as well as other relevant documents, at a later point.

### 6.1.1 Research Questions

Grossmann's contribution to the genesis of GR has two aspects that we will carefully distinguish. On the one hand, there are Grossmann's *passive contributions* to GR. At some point in 1912, Einstein asked Grossmann for help with the existing mathematical literature. This means that one part of Grossmann's job was simply to scan the mathematical literature, and show and explain the results of his search to Einstein. This task requires mathematical knowledge, but no substantive, original mathematical contribution on Grossmann's part.

If we want to understand this aspect of Grossmann's contribution, we will have to dig into the mathematical literature that Grossmann used, in order to understand the existing mathematical knowledge. More specifically, we will answer the following questions:

- What are the mathematical sources that Grossmann used for his contribution to GR?
- What are the mathematical theories and traditions behind these sources?
- What is the role, and relative importance, of these theories and traditions (in Grossmann's eyes) for the mathematical development of GR?

On the other hand, we want to gauge the extent of Grossmann's *active contributions* to GR. We will analyze whether Grossmann had to modify the existing mathematical theories, and whether he contributed new, original pieces of mathematics to GR. In particular we will answer the following questions:

- To what extent did Grossmann transform, amend and simplify the available mathematical results for the purposes of application in GR?
- Did Grossmann contribute genuinely original mathematical results?
- What is the relative importance of Grossmann's new results?

This set of questions is not independent of the first, as we need a firm grasp on the existing mathematical knowledge to gauge the scope of Grossmann's own contributions. We thus face the task of surveying the mathematical knowledge as completely as possible. In the present study, we will only take into account the mathematical sources that Grossmann cites; the task of tracking his influences thereby becomes tractable. A short overview of the most important mathematical sources can be found in chapter 5. We will also draw on selected secondary sources, on relevant mathematical theories and concepts, in order to embed the primary sources in their respective mathematical context.

### 6.1.2 Method

One of our tasks is to identify the origin of the mathematical theories and notions presented by Grossmann. In a first step, we will compare Grossmann's account with its various predecessors. However, it may happen that more than one of the sources discusses some particular result. In this case, we will compare Grossmann's notation with that of the sources in question. It is plausible that, whenever Grossmann's notation is very close to the notation of one of the sources, he mainly used this source for a particular concept or theory. Finally, we will also use the manner of citation, to track the lines of influence. We use historical notation throughout the chapter.

Our use of notation for identifying sources has to be taken with a grain of salt. In general, the more a mathematician is familiar with some mathematical theory, the more his notation will be independent of the source he uses, i.e. the link between notation and sources is weakened. Thus, our method presupposes, to a certain degree, that Grossmann was not an expert on the ADC. Our examination of the Entwurf, and on his scientific biography, suggests that this is the case, and that Grossmann followed the sources rather closely.

We are also interested in Grossmann's original contributions. Here we will, to a certain extent, rely on secondary sources. If some result cannot be traced to any of Grossmann's sources, we will tentatively attribute it to him. Certain results have been discovered, and discussed, before, without Grossmann's acknowledgement. Some of these results may have been rediscovered by Grossmann. This is plausible, especially if the context of the result is far from Grossmann's interests. We are not interested in Grossmann's originality, but in the extent of his active and passive contributions.

### 6.1.3 Overview

Here is an overview of the chapter. In section 6.2, we give a short summary of Grossmann's mathematical sources, and list the citations, illustrated with a citation tree. In section 6.3, we discuss one of the sources that is often taken to be particularly important and influential: Riemann's contributions, which outline a geometry with variable curvature. We argue that Grossmann may have never consulted Riemann's work. In section 6.4, we examine the introduction to Grossmann's part, focusing on his view of the existing mathematical literature, his own contribution, and his methodological credo. In 6.5, we take a closer look at the fundamental concept of manifold, in the Entwurf and the mathematical literature. In sections 6.6 to 6.9, we continue or examination of the Entwurf, tracing the origin of the mathematics at each step. We pay special attention to the origin of the mathematics in the (failed) generally covariant approach to the field equations. Finally, in 6.10, we summarize our results and note systematic consequences and open questions.

## 6.2 Grossmann's Sources

In this section, we describe and analyze the citations of works of pure mathematics in Grossmann's part of the Entwurf. We add a citation tree, which shows the citations by Grossmann and by his sources. Before we examine the citations, a very brief characterization of the main protagonists may be helpful; see also the discussions in chapter 5.

Grossmann cites Riemann, Christoffel, Ricci & Levi-Cività, Kottler, and Bianchi-Lukat.

**Riemann:** Two of his contributions are potentially relevant. His paper on the foundations of geometry, which is based on his habilitation lecture, formulates seminal concepts such as manifolds, and the idea of a geometry of variable curvature. This paper is informal in style, as it is aimed at a general audience. The "Commentatio", on the other hand, is a paper on the heat equation, which introduces and discusses the eponymous Riemann tensor.

**Christoffel's** paper is a contribution to algebraic invariant theory, in an algebraic, algorithmic tradition. He aims to solve a technical problem, the equivalence problem of homogeneous quadratic differential forms, and is not interested in geometry.

**Ricci & Levi-Cività's** famous "tensor analysis paper" presents a general calculus, the ADC, and demonstrates its applicability in geometry, analysis, and physics. The pure calculus is heavily indebted to Christoffel. The paper has a survey character, and proofs are often omitted.

**Bianchi-Lukat** is the German translation of an Italian textbook on differential geometry. Grossmann mainly used the second chapter, which also draws on Christoffel's work. However, Bianch's presentation is more accessible than Christoffel's.

**Kottler** applies the ADC to a physical problem. This is no pure mathematics paper; however, it contains some relevant mathematical results. This is one of the first physics papers to apply the ADC. Grossmann's motivation for citing it could be priority: Grossmann has a an alternative proof of a result by Kottler.

Here are the citations of these sources in full (references in square brackets are to (Klein et al., 1995)):

- "Christoffel: Über die Transformation der homogenen Differentialausdrücke zweiten Grades, J. f. Math. 70 (1869), S. 46." [p. 324]
- "Ricci et Levi-Cività, Méthode de calcul différentiel absolu et leurs applications, Math. Ann 54. (1901), S. 125." [p. 324]
- "Kottler, Über die Raumzeitlinien der Minkowskischen Welt, Wien. Ber. 121 (1912)." [p. 324]
- "Bianchi-Lukat, Vorlesungen über Differentialgeometrie, erste Auflage, S. 47.", [p. 330]
- "Riemann, Ges. Werke, S. 270." [p. 336]

These are the primary mathematical sources that Grossmann used. From a close examination of the manner and place of these citations, we can gain valuable information about these sources, their relative importance, and the relations between these sources.

The first three sources, Christoffel, Ricci & Levi-Cività, and Kottler, are first cited in the introduction; Bianchi-Lukat and Riemann in the main text. Grossmann further mentions "Minkowski, Sommerfeld, Laue u.a." as contributors to the vector-analytic innovations in special relativity. On p. 328, he cites the relevant papers of these three authors. Other mathematicians, including Laplace and Beltrami, are mentioned without citation. Laplace is mentioned in the context of the Laplace operator, Beltrami in the context of Beltrami parameters.

Some of the sources are mentioned more than once. Here are the further citations and their context:

**Ricci & Levi-Cività:** p. 326, footnote 1 – remarks on their tensor notation and why Grossmann's deviates; p. 329 – attribution of the name "covariant differentiation"; p. 333 – attribution of the discriminant tensor ("system $\epsilon$").

**Christoffel:** p. 328 – attribution of the form of covariant differentiation; p. 329 – introduction of the Christoffel symbols of the first and second kind: p. 336 – attribution of the Riemann tensor.

**Kottler:** p. 331 – attribution of one form of the divergence of a covariant four-vector.

The fact that some of the sources are cited more than once gives us valuable information about their importance. It is plausible that the mathematical sources with multiple citations were really consulted and used by Grossmann, especially if the reference is not contained in one of the other sources; we will see in the next section that this is relevant in Riemann's case. Based on the number of citations, the most important mathematical sources are Ricci & Levi-Cività, and Christoffel, followed by Kottler, and finally Riemann and Bianchi-Lukat.

### 6.2.1 Citation Tree

The citation tree shows who cites whom among Grossmann's sources. We will return to the significance of this network of citations in the course of the chapter.

## 6.3 The Curious Case of Riemann

Grossmann's citation of Riemann's habilitation paper is of great interest. Riemann is commonly taken to have had a big influence on the genesis of GR: he discovered geometries of variable curvature in $n$ dimensions, and in particular the eponymous Riemann tensor[2], one of the crucial mathematical objects of GR. However, we think that, based on Grossmann's citation, he did not actually consult Riemann's work for the Entwurf. Here are our reasons for this claim.

Grossmann cites Riemann one time, on p. 336: "Riemann, Ges. Werke, S. 270.". First, this reference is faulty. There are two editions of Riemann's collected work to which Grossmann had access, Riemann (1876b) and Riemann (1892). The reference to p. 270 does not make sense for both editions. In the 1876 edition, page 270 is the beginning of "Ein Beitrag zur Elektrodynamik". In the 1892 edition, page 270 is the second-to-last page of "Ueber die Darstellbarkeit einer Function durch eine trigonometrische Reihe". Both papers are irrelevant in the present context. Note, however, that in the 1892 edition, Riemann's habilitation paper "Ueber die Hypothesen, welche der Geometrie zu Grunde liegen" begins on p. 272.

[2]Einstein calls it the Riemann-Christoffel tensor. This is a suggestive name, not because it hints at questions of priority, but because it indicates that the tensor has its roots in at least two different mathematical traditions, one more geometrical, the other invariant-theoretic in nature.

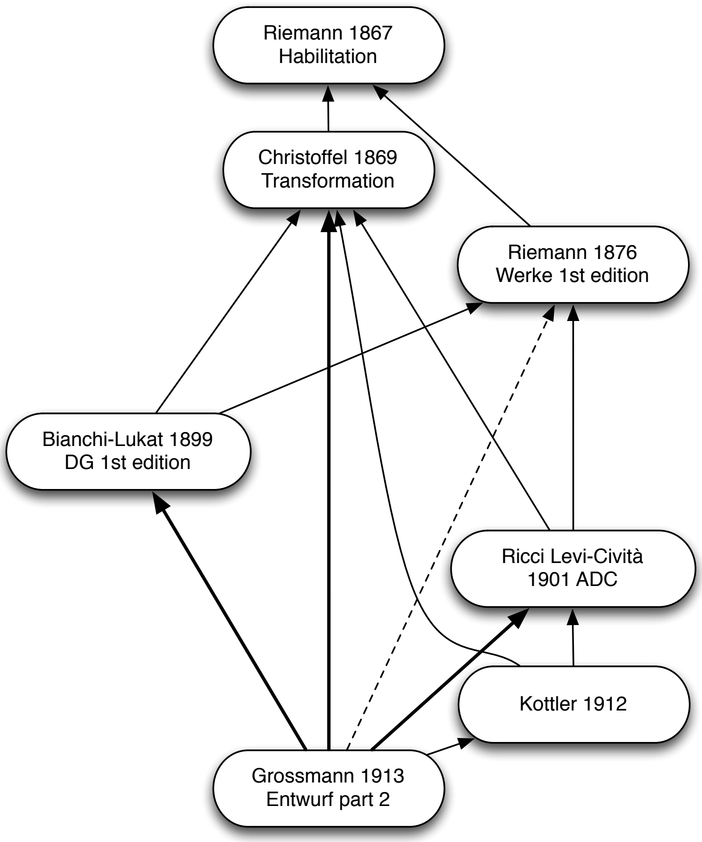



Figure 6.1: Grossmann's Sources: Citation Tree

Grossmann's mistake has been pointed out in Janssen et al. (2007b, p. 611, fn. 209). The authors write: " '270' is a misprint and should be '370' " – in the 1876 edition, the "Commentatio", in which the Riemann tensor is discussed, begins on that page. If this were indeed so, Grossmann's mistake would be an uninteresting typo.[3]

We agree that this is one possibility. However, there is an even better explanation. We believe that Grossmann's mistake is not a typo, but rather that he copied a typo from Ricci and Levi-Civita (1901). It is well known that Grossmann made extensive use of the ADC paper. Ricci & Levi-Cività cite Riemann's works twice:

- On p. 142, they introduce the "covariant system of Riemann"[4], i.e. the

[3]In a footnote to Grossmann's reference to Riemann in Klein et al. (1995), we are referred to Riemann (1892). In view of the above, the reference should be to Riemann (1876b, p. 370).

[4]"système covariant de Riemann"

Riemann tensor. They write that "[the symbols] can be found in the *Commentatio mathematica* by Riemann" ("On les trouve dans la *Commentatio mathematica* de Riemann."), adding the following reference in a footnote: "Gesammelte Werke, pag. 270".

- On p. 192, they discuss physical application of the ADC and mention that Riemann solved a certain problem of the heat propagation equation, adding a reference in a footnote:" 'Commentatio mathematica, qua etc.', Ges. Werke, pag. 370".

In the 1876 edition, the "Commentatio" begins on p. 370, while in the 1892 edition, the "Commentatio" begins on p. 391. Thus, the first citation by Ricci & Levi-Cività on p. 142 is wrong, it should be to p. 370 of the 1876 edition. What is noteworthy about this typo is that it is identical to Grossmann's – both citations are identically wrong. The only difference between the two is that Grossmann writes "S. 270", whereas Ricci & Levi-Cività write "pag. 270". It is thus reasonable to conjecture that Grossmann simply copied the faulty citation from Ricci & Levi-Cività. It would be a huge coincidence if both Grossmann and Ricci & Levi-Cività had independently made the same typo when citing Riemann. What is more, the first passage in which Ricci & Levi-Cività cite Riemann is central for Grossmann, because it is here that they introduce the Riemann tensor. Ricci & Levi-Cività's second (correct) citation of Riemann, on the other hand, is not directly relevant for Grossmann – it is about an application of the ADC.

The fact that Grossmann copied Ricci & Levi-Cività's mistake has an interesting consequence. Recall that this is the only time Grossmann even mentions Riemann. It is therefore probable that Grossmann did not actually consult Riemann's work for the mathematical part of the Entwurf, which, in turn, suggests that Riemann's work had no direct influence on the very first tensorial formulation of GR. This applies not only to Grossmann, but also to Einstein. Einstein probably first learned about the Riemann tensor from Grossmann; this is documented in the Zurich notebook, p. 14L (see Janssen et al. (2007a, p. 418)), where Einstein writes down the Riemann tensor, with Grossmann's name next to it.

Some remarks in Janssen et al. (2007b, pp. 610) suggest that Riemann might have had a direct influence on Grossmann nevertheless. The authors discuss where Grossmann might have learned about the Riemann tensor. They mention that Christoffel and Riemann are candidates, and give two reasons as to why Riemann might have had a direct influence on Grossmann and Einstein.

The first reason is a notational detail, the use of a comma in the symbol for the Riemann tensor, which was used by Riemann, but not by Christoffel. However, Riemann is not the only one to use this comma. It also features in Bianchi and Lukat (1899). We will discuss the notational variants in section 6.9.2 below.

The second reason why there might have been a direct influence by Riemann on Grossmann is the occurrence of the word "Mannigfaltigkeit", manifold, which features in Riemann's work, but not in Christoffel's paper. We agree. However, the manifold concept features, under its French name ("variété $V_n$"), in Ricci and Levi-Civita (1901), in prominent positions, such as in the introduction. We will discuss the interpretation of Grossmann's concept of manifold in detail in section 6.5 below.

Thus, both reasons mentioned by Janssen et al. for thinking that Riemann had a direct influence on Grossmann, can be explained by taking other mathematical sources into account.

The result that Grossmann did not consult Riemann directly has to be taken with a grain of salt. We do not deny, in any way, that Riemann's work was important for the genesis of GR. Riemann's influence shows up in all of the mathematical sources used by Grossmann. However, the fact that Riemann's influence is mediated by others is nevertheless important, because some of the mathematicians building on Riemann had quite a different methodology, and perspective, on the relevant mathematical techniques and results. We will have to keep this in mind while we track Grossmann's sources.

## 6.4 Introduction to Part II

In this section, we begin our analysis of Grossmann's part of the Entwurf. The introduction to part II is of great interest for our purposes. Here Grossmann states which mathematicians and mathematical theories he draws on; furthermore, some passages indicate what he took to be his contribution to the mathematics of GR. We first give a translation of the introduction, then an analysis.

> The mathematical tools necessary to devise a vector analysis of gravitational fields characterized by the invariance of the line element
>
> $$ds^2 = \sum_{\mu\nu} g_{\mu\nu} dx_\mu dx_\nu \qquad (6.1)$$
>
> can be traced back to the fundamental treatise by *Christoffel* [...] on the transformation of quadratic differential forms. Based on *Christoffel's* results, *Ricci* and *Levi-Cività* [...] have devised their methods of an absolute differential calculus, that is, a calculus independent of the coordinate system; these methods make it possible to state the differential equations of physics in an invariant manner. However, as the vector analysis of Euclidean

> space in arbitrary curvilinear coordinates is formally identical to the vector analysis of an arbitrary manifold given by its line element, it is not difficult to extend the vector-analytic concepts ("Begriffsbildungen") that have been developed in recent years by *Minkowski, Sommerfeld, Laue et al.* for relativity theory, to the preceding general theory by *Einstein*.
>
> With some practice, the *general vector analysis* thus obtained can be manipulated as easily as the special vector analysis of the three- or four-dimensional euclidean space; indeed, the greater generality of its concepts ("Begriffsbildungen") lead to a clear exposition that cannot be found in the special case.
>
> The theory of special tensors (§ 3) has been thoroughly covered in a treatise by *Kottler*, which appeared while the present paper was being written; Kottler's work is based on the theory of integral forms, an approach that is not generalizable.
>
> A systematic presentation of general vector analysis may be appropriate, as a more thorough mathematical study will have to follow the gravitational theory by *Einstein* and especially the problem of the differential equations of the gravitational field. In doing this I did not draw on geometrical tools, as they contribute little to the illustration ("Veranschaulichung") of the concepts ("Begriffsbildung") of vector analysis.

In the first sentence of the introduction, Grossmann characterizes his goal in general terms: the development of a vector analysis based on the invariance of the line element. The tools for this vector analysis have their origin in Christoffel's work. The fact that Grossmann mentions Christoffel in the very first sentence, and calls his paper "fundamental", underlines the importance of Christoffel's work, as does the characterization of the mathematical problem, that of the *invariance* of the line element. In the second sentence, Grossmann notes that Ricci & Levi-Cività have worked out Christoffel's results and devised methods for stating differential equations in an invariant manner; he emphasizes the coordinate independence of their theory.

These first two sentences introduce what Grossmann considers to be his most important mathematical predecessors: Christoffel on the one hand, and Ricci & Levi-Cività on the other. The next sentence is already about Grossmann's own contribution.

The characterization of Christoffel as the "main technical innovator" of the ADC is probably not Grossmann's; it can be found in the introduction of Ricci & Levi-Cività's paper. They write: "The algorithm of the absolute differential Calculus, that is the material instrument of the methods ("l'instrument matériel des méthodes") to which we will introduce the readers [...] can be found entirely in a remark due to M. Christoffel [...]" (Ricci and Levi-Civita, 1901, p. 127). Here the "material instrument" should be contrasted with what the calculus describes, n-dimensional manifolds, which goes back to Gauss and Riemann. It seems that Grossmann adopted Ricci & Levi-Cività's perspective on the division of labour between their own contribution and Christoffel's, without mentioning Gauss and Riemann.[5]

The third sentence describes Grossmann's own contribution: the connection between the "vector analysis of an arbitrary manifold given by its line element", i.e. the mathematics of the ADC, and the traditional mathematics of "vector analysis" as it is known from "relativity theory", i.e. SR. This connection is established via the "vector analysis of the euclidean space, related to arbitrary curvilinear coordinates": Grossmann conceives of this vector analysis as a generalization of the vector analysis of SR, which is limited to pseudo-Euclidean coordinates. The distinction between this generalization and the vector analysis of a manifold is a formal one.

According to Norton (1992, Appendix), Grossmann is the first to systematically bring together the two traditions of vector analysis, as developed in physical application, and the ADC. In the Entwurf, Grossmann draws this distinction in the abstract; later, in the Frauenfeld lecture, he explains the parallels in more detail.

The second paragraph of the introduction advertises the advantages of the new mathematical approach: the new theory is easy to use and clearly arranged. The third paragraph mentions work by Kottler on the theory of special, i.e. antisymmetric, tensors. Grossmann points out that, while Kottler's treatment of special tensors is "complete", it is not based on the general vector analysis he will draw on, but on the theory of integral forms. In the last paragraph, Grossmann appears to acknowledge that he is not completely satisfied with his results, or the manner in which they are presented. On the other hand, it is a justification for the rather detailed exhibition of vector analysis that is to follow.

Grossmann's remarks on geometry could be somewhat puzzling for the modern reader, as they appear to be off the mark. Grossmann did not anticipate the fundamental importance of geometrical notions for GR. The only role he envisioned for geometry was that of an illustration, and he considered geometry to be of rather limited usefulness, even in this respect. His remarks are, however, very much in line with Ricci & Levi-Cività's ideas. In their introduction, they emphasize the value of the ADC's abstractness; what is more, they appear to find it distasteful if a certain "purity of method" is not met, e.g. the use of variational methods to derive Beltrami parameters.[6]

[5]Karin Reich (1994, p. 81) points out that Ricci & Levi-Cività trace the ADC back to Beltrami's work on differential parameters. It seems to us that the reason why Ricci & Levi-Cività mention Beltrami is that, historically, grappling with Beltrami's work led to the ADC. Their goal was to find the proper mathematical framework for Beltrami's results. We will return to this point below.

[6]Dell'Aglio (1996) traces this emphasis on the tradition of algebraic invariants back to the great influence of Christoffel on Ricci's work.

This call for a purity of methods was probably heard by Grossmann.

## 6.5 Interlude: Manifolds in the Entwurf

In this section, we discuss Grossmann's manifold concept in the Entwurf, as well as its historical origins.[7] Manifolds are an indispensable tool for the mathematical formulation of GR, and the concept has an involved history in this context – think of the Hole Argument.

In the previous section, we mentioned the thesis that Riemann's work was important for the first tensorial formulation of GR, because manifolds feature in the mathematical part of the Entwurf. Janssen et al. (2007b, pp. 610) interpret the use of the notion as a sign that Riemann had a direct influence on Grossmann. We argued that this is not necessarily the case, as Grossmann might have adopted the notion from Ricci & Levi-Cività. This suggests that, while Riemann influenced Grossmann, the influence is indirect and filtered through Ricci & Levi-Cività's lenses. We will examine Ricci & Levi-Cività's conception in some detail. Other possible interpretations of how Einstein and Grossmann conceived of manifolds emerge from a search for the word "manifold" in Einstein's other writings.

### 6.5.1 Very Early History

The concept of manifold goes at least back to Gauss, who uses the concept, albeit cautiously, in the "disquisitiones"; see section 5.2.3 for Gauss's theory of two-dimensional manifolds. The modern notion has its origin in the writings of Riemann, especially his habilitation paper (see section 5.2.3). Riemann already had an elaborate manifold concept that was important for his work on differential geometry, as well as other subfields, such as complex analysis and topology. For our purposes, his treatment of continuous, "metric" manifolds, i.e. metrics with more than just topological structure, is relevant. In his habilitation paper, he describes the line element, and how it determines lengths in such manifolds, without using formulas. The manifold "fully determines" the $n(n-1)/2$ functions of the metric that are not determined by the $n$ variable transformations.

After Riemann, several mathematicians used manifolds as a tool in the foundations of geometry.[8] Exploring the influence of these mathematicians on Grossmann and Einstein is beyond the scope of this chapter; a few remarks will have to suffice. Two of the authors who developed the notion further are Beltrami and Klein. There are paths from both of these authors to Grossmann.

[7]The historical discussion of manifolds draws on Scholz (1980). We thank Erhard Scholz for correspondence concerning this section.

[8]See Scholz (1980, ch. 3)

Beltrami is relevant for Grossmann because, for one, he worked on non-euclidean geometry, in particular surfaces of constant, non-zero curvature; a topic that was of great interest to Grossmann. Also, Beltrami was a strong influence on Ricci. Klein, on the other hand, certainly had an influence on Grossmann via Bianchi. Bianchi studied in Göttingen for two years and was influenced by Klein.[9] Grossmann consulted Bianchi's textbook when he worked on the Entwurf; see section 6.7 below. The textbook is also relevant in the context of surfaces with constant curvature. We also know that Grossmann carefully studied Klein's work on projective geometry, since he worked on Cayley-Klein metrics.

### 6.5.2 Ricci & Levi-Cività

Norton (1992, p. 308) argues that Ricci and Levi-Civita (1901) set a precedent for Grossmann's perspective on manifolds:

> Ricci and Levi-Civita buried their definition of a manifold in the short preface within an account of the geometric ancestry of their absolute differential calculus. The formal exposition of their calculus begins in Chapter 1, with no mention of manifolds and in a way that seems to seek as much of a divorce from geometrical associations as possible.

This approach, Norton thinks, carries over to Grossmann: "[T]he concepts of manifold and coordinate system were to be taken as terms already known to the reader. At best, they were to be dismissed briefly in prefatory remarks." (Ibid.)

We agree that Ricci & Levi-Cività could be the origin of Grossmann's manifold concept. However, we do not agree with Norton's assessment of the importance of manifolds for Ricci & Levi-Cività. First of all, the discussion of manifolds takes up a good part of the introduction, which underlines the importance of the concept. More importantly, their discussion does not suggest that manifolds are only important for the "geometric ancestry" of the more general ADC. True, the origin of the ADC is traced to a geometric tradition as opposed to its "algorithmic core", which goes back to Christoffel. However, Ricci & Levi-Cività think there is an intimate connection between the ADC and manifolds (p. 127f.):

> A manifold $V_n$ is defined intrinsically in its metrical properties by $n$ independent variables and by a whole class of quadratic differential forms of these variables, any two of which can be transformed into each other by a point transformation. As a

[9] See Reich (1989, p. 282).

> consequence, the $V_n$ [the manifold] is invariant under all transformations of its coordinates. The absolute differential calculus, by acting on the covariant and contravariant differential forms of $ds^2$ [the line element] of $V_n$ [the manifold] in order to derive others of the same nature, is also, in its formulae and in its results, independent of the choice of coordinates. – Being of the sort essentially attached to $V_n$ [the manifold], it is the natural instrument of all investigations which have such a manifold as their object, or in which one encounters as characteristic element a positive, quadratic form of the differentials of $n$ variables or their derivatives.[10]

Especially in the last sentence, Ricci & Levi-Cività propose a close, "natural" connection between the ADC and manifolds: They characterize the ADC as being "essentially attached" to a manifold and as being the natural tool to investigate such objects. Manifolds are always implied when one uses the ADC – it is what the calculus refers to.

The link between the line element and the manifold emerging from this quote is striking. According to the last sentence, the line element "belongs to" a manifold, or can stand in for a manifold. This might explain why manifolds are not mentioned in the exposition of the ADC as a calculus, a fact pointed out by Norton: if the line element is nothing but an algebraic representation of the manifold, there is no need to deal with manifolds directly. We will see below that this link between manifold and line element is key to Grossmann's conception as well.

The importance of the concept of manifolds shows not only in the introduction, but also throughout Ricci & Levi-Cività's paper. For example, manifolds feature quite prominently in the second part of the paper on intrinsic geometry; see pp. 145. Here is the beginning of the chapter:

> In this chapter, we will draw on the language of geometry by considering the fundamental form $\phi$ to be the $ds^2$ of a manifold $V_n$.[11]

[10]"[U]ne variété $V_n$ est définie intrinsèquement dans ses propriétés métriques par $n$ variables indépendantes et par toute une classe de formes quadratiques des différentielles de ces variables, dont deux quelconques sont transformables l'une en l'autre par une transformation ponctuelle. – Par conséquence une $V_n$ reste invariée vis-à-vis de toute transformation de ses coordonnées. Le Calcul différentiel absolu, en agissant sur des formes covariantes ou contrevariantes au $ds^2$ de $V_n$ pour en dériver d'autres de même nature, est lui aussi dans ses formules et dans ses résultats indépendant du choix des variables indépendantes. – Étant de la sorte essentiellement attaché à $V_n$, il est l'instrument naturel de toutes les recherches, qui ont pour objet une telle variété, ou dans lesquelles on rencontre comme élément caractéristique une forme quadratique positive des différentielles de $n$ variables ou de leurs dérivées."

[11]"Dans ce chapitre nous aurons recours au langage géométrique en considérant la forme fondamentale $\phi$ comme le $ds^2$ d'une variété $V_n$ à $n$ dimensions."

Manifolds are treated as general, n-dimensional geometrical objects throughout the chapter.[12]

### 6.5.3 Grossmann

The assumption that Grossmann adopted Ricci & Levi-Cività's perspective on manifolds has interesting consequences. It would not only explain why manifolds are almost constantly associated with the line element, but also why manifolds do not feature prominently in the Entwurf. As we saw in the discussion of the introduction to the Entwurf, Grossmann refrains from using "geometrical tools" in the exposition of tensor calculus. If he had a geometric conception of manifolds, it is reasonable that they do not feature later in the Entwurf.

An analysis of the entire Entwurf reveals that manifolds only feature in Grossmann's part – Einstein never uses the word "manifold". In Grossmann's part, manifolds feature in the introduction, in part three (once), and then in part four, section two; the discussion of the (failed) attempt to find generally covariant field equations.

As we already noted, Grossmann uses the concept of manifolds in almost constant association with the line element: a manifold is "given by its line element". This formulation is used in the introduction as well as in the discussion; in part four, section two. In part three, Grossmann even identifies the manifold with the line element: he introduces the Levi-Cività symbol and emphasizes its "importance for the vector analysis of the n-dimensional manifold $ds^2 = \sum_{\mu\nu} g_{\mu\nu} dx_\mu dx_\nu$". A very close association indeed.

The close connection between manifold and line element suggests that Grossmann, just as Ricci & Levi-Cività, regarded line element and metric as an (algebraic) surrogates that facilitate calculations without invoking dubious, geometrical connotations. The line element becomes the central object and makes geometric reasoning superfluous.

However, maybe we modern readers are simply reading too much into the notion, as we are aware of the conceptual difficulties that arose in the genesis of GR, from the question of how to interpret manifolds. Our reading could be anachronistic indeed, as a glance at Einstein's use of the concept shows. A search for the word "manifold" in Einstein's writings reveals a more colloquial, non-technical use of the word, which is derived from the German adjective "mannigfaltig", meaning "diverse". Maybe manifolds are nothing but an undetermined "diversity", or a stand-in for whatever object

[12]We have to point out that the importance of manifolds cannot be adequately assessed if one consults Hermann (1975); the standard english translation of Ricci & Levi-Cività's paper. Hermann's translation almost systematically suppresses the notion of manifold. For example, the above quotation reads as follows in Hermann's translation: "In this chapter we make use of geometric language, with a Riemannian metric tensor $\phi$ defining the basic 'geometry.' [sic] " (ibid., p. 65). The translation does not live up to historical standards in other respects as well.

is represented by the line element. Maybe Einstein and Grossmann did not pay too much attention to the concept, because they thought that the relationship between metric and manifold is straightforward, or even trivial, or because they worked under time pressure and pressed on without working out the ramifications of the concept.

Finally, we have indirect evidence that Grossmann may have emphasized the algebraic, invariant-theoretic perspective of Ricci & Levi-Cività, originating in Christoffel, over Riemann's more geometric approach, from a letter that Felix Klein wrote to Einstein on March 20, 1918:

> You will probably immediately agree with what I have to say about Riemann, Beltrami, and Lipschitz; it seems to me that Grossmann at the time instructed you too much from the point of view of the school of Christoffel more narrowly.[13]

Summing up, we have various possible interpretations of the concept of manifold in the Entwurf, which are not necessarily mutually exclusive. First, Grossmann could have adopted a more or less sophisticated version of the concept, as developed in the mathematical literature, especially by Riemann. We think that a direct influence of Riemann is unlikely, as the discussion in the Entwurf is not very deep, and Grossmann may not have consulted Riemann. Second, Grossmann could have relied on Ricci & Levi-Cività's treatment of manifold. This is Norton's preferred account. Grossmann could have adopted Ricci & Levi-Cività's separation of the algorithmic aspect of the ADC, which mainly deals with the line element, and its natural interpretation in terms of geometry, which involves the manifold. Third, there is a more extreme reading of Grossmann's interpretation of the relation between manifold and line element, according to which the two concepts are to be identified. Fourth, the use of "manifold" may have been influenced by the colloquial meaning of the word at the time. This last interpretation cautions us against reading too much of the modern notion into the concept used in the Entwurf.

## 6.6 Paragraph 1: General Tensors

In this paragraph, Grossmann introduces tensors, and what is now known as tensor algebra; operations allowing the construction of new tensors from old ones. We have divided the paragraph into subparagraphs, in order to enhance readability.

[13]"Was ich von Riemann, Beltrami, und Lipschitz erzähle, wird wohl gleich ihren Beifall haben; es scheint mir, dass Grossmann Sie s. Z. zu einseitig vom Standpunkte der engeren Christoffelschen Schule aus instruiert hat" – cited after (Janssen et al., 2007b, p. 611, fn. 212).

### 6.6.1 Variables and their Differentials, Fundamental Tensor and its Inverse, Transformation Properties

Grossmann first introduces the square of the general line element, as in equation (6.1) above. He writes that it "can be interpreted as the invariant measure of the distance of two infinitely-close space-time points". This interpretation is new, insofar as the ADC has not been applied to theories of space-time before. However, the exposition of tensors is largely independent of this interpretation; accordingly, all concepts are defined for $n$ variables.

Next, Grossmann discusses transformations. If one set of variables is a general function of another set of variables, $x_i = x_i(x'_1, x'_2, ..., x'_n)$, their differentials transform as

$$dx_i = \sum_k \frac{\partial x_i}{\partial x'_k} dx'_k = \sum_k p_{ik} dx'_k \tag{6.2}$$
$$dx'_i = \sum_k \frac{\partial x'_i}{\partial x_k} dx_k = \sum_k \pi_{ki} dx'_k$$

The notation $p_{ik}$, $\pi_{ik}$ is not used by Ricci & Levi-Civià in this form, nor by Christoffel, who uses a more compact notation. Bianchi (p. 38) uses the notation $dx_r = \sum_i p_{ri} dx'_i$ with $p_{ri} = \frac{\partial x_r}{\partial x'_i}$ for variable differentials. Grossmann then notes the transformation behavior of the line element $g_{\mu\nu}$, using the same notation.

Next, Grossmann introduces the discriminant of the fundamental tensor, the determinant $g = |g_{\mu\nu}|$. This allows him to define $\gamma_{\mu\nu}$, a quantity he only later calls the inverse fundamental tensor; here it is defined, rather awkwardly, as the "normed subdeterminant of $G$ adjoint to $\gamma_{\mu\nu}$" ("$\gamma_{\mu\nu}$ [ist] die durch die Diskriminante dividierte ('normierte'), dem Element $g_{\mu\nu}$ adjungierte Unterdeterminante von $g$"). The $\gamma_{\mu\nu}$ transform as

$$\gamma'_{rs} = \sum_{\mu\nu} \pi_{\mu r} \pi_{\nu s} \gamma_{\mu\nu} \tag{6.3}$$

The manner in which $\gamma_{rs}$ is introduced is telling. The notation for the components of the metric, and especially for the inverse, is, in all probability, not from Ricci & Levi-Civià; on p. 134, they introduce the inverse metric as such, and show how covariant and contravariant tensors ("systèmes") can be transformed into each other. The discriminant is not used in the definition of the inverse metric.

Christoffel uses the subdeterminant definition, in what is, essentially, the definition of the inverse of the metric, but he does not explicitly address the relation between covariant and contravariant quantities.

Grossmann's presentation is closest to Bianchi's. Bianchi introduces the components of the inverse metric as $A_{ks}$ ($= \gamma_{ks}$), on p. 37; he uses the same definition as Grossmann ("$A_{ks}$ [... ist] die durch die Discriminante $a$ selbst dividierte Unterdeterminante von $a_{ks}$ in $a$"). This is the second instance where Grossmann follows Bianchi rather than Ricci & Levi-Cività.

### 6.6.2 Covariant, Contravariant and Mixed Tensors

Grossmann defines general covariant, contravariant, and mixed tensors, based on their transformation behavior. Covariant tensors are written with latin capitals, contravariant tensors with Greek capitals, and mixed tensors with German capitals:

$$\begin{aligned}
T'_{r_1 r_2 \ldots r_\lambda} &= \sum_{i_1, i_2 \ldots i_l} p_{i_1 r_1} p_{i_2 r_2} \ldots p_{i_\lambda r_\lambda} \cdot T_{i_1 i_2 .. i_\lambda} \\
\Theta'_{r_1 r_2 \ldots r_\lambda} &= \sum_{i_1, i_2 \ldots i_l} \pi_{i_1 r_1} \pi_{i_2 r_2} \ldots \pi_{i_\lambda r_\lambda} \cdot \Theta_{i_1 i_2 .. i_\lambda} \\
\mathcal{T}'_{r_1 r_2 \ldots r_\mu / s_1 s_2 \ldots s_\nu} &= \sum_{\substack{i_1, i_2 \ldots i_\mu \\ k_1, k_2 \ldots k_\nu}} p_{i_1 r_1} p_{i_2 r_2} \ldots p_{i_\mu r_\mu} \cdot \pi_{k_1 s_1} \pi_{k_2 s_2} \ldots \pi_{k_\nu s_\nu} \cdot \mathcal{T}_{i_1 i_2 \ldots i_\mu / k_1 k_2 \ldots k_\nu}
\end{aligned}$$

Consequently, he classifies $g_{\mu\nu}$ as a covariant tensor of order two, $\gamma_{\mu\nu}$ as a contravariant tensor of order two, and variable differentials as contravariant tensors of order one. In the case $n = 4$, $g_{\mu\nu}$ and $\gamma_{\mu\nu}$ are interpreted as the fundamental tensors of the gravitational field.

In a footnote to the contravariant case, Grossmann comments on the definition and notation of tensors. He notes that "covariant (contravariant) tensors of rank $\lambda$ are thus identical to the 'covariant (contravariant) systems of degree $\lambda$' by Ricci & Levi-Cività". However, he does not want to adopt their notation, as "complications with compounded equations have forced us to choose the above notations".

Grossmann is the first to introduce the notion of tensor for these mathematical objects, i.e. for general co- and contravariant quantities of arbitrary rank.[14] Ricci & Levi-Cività call these objects "systems". Previously, only special objects with a physical interpretation from, e.g., elasticity theory were called tensors. It is also the first definition of general mixed tensors.[15]

It is not entirely clear why Grossmann chose the more awkward notation, with different kinds of letters, over the more elegant system, with upper and lower indices, proposed by Ricci & Levi-Cività; a notation that is still used today. One possible explanation is the use of mixed tensors in Einstein's part, especially the parts on electrodynamics.[16] Also, a mixed tensor, for

[14] See the appendix of Norton (1992), and especially the Addendum on pp. 309, for more on Grossmann's use of the notion.

[15] See Reich (1994, p. 194)

[16] Abraham Pais (1982, p. 220) notes that the Entwurf contains the correct generally covariant version of the Maxwell equations.

the divergence of the stress-energy tensor, features in Grossmann's part of the Frauenfeld lecture.

In sum, in this subsection we have first examples of new concepts and notation. We do not yet fully understand the motivations behind these innovations; they could well be motivated by requirements from existing physical theories.

### 6.6.3 Tensor Algebraic Operations

Grossmann introduces tensor-algebraic operations: sum; outer product, now called tensor product; inner product, including the general case of mixed tensors; "reciprocity", now called raising and lowering indices, and; contracting indices.

The end of the paragraph has a short discussion of how invariants can be formed from co- and contravariant tensors of order one and two, and dimension 4, by using the metric. Grossmann notes the form of these invariants in SR, but then states that he will no longer explore the parallel to SR; instead, he refers the reader to Minkowski, Sommerfeld and Laue.

The tensor algebraic operations are very similar to Ricci & Levi-Cività's, up to their names; for example "Reziprozität" corresponds to "*réciproques* par rapport à la forme fondamentale". There are, however, some notational deviations; also, the comparison to SR, and the introduction of the operations for mixed tensors, is new.

## 6.7 Paragraph 2: Differential Operators on Tensors

In the second paragraph, Grossmann introduces tensor calculus proper: "extension" ("Erweiterung"), the covariant derivative, divergence, and "generalized Laplacean operation"; Laplace-Beltrami differential operators. Grossmann then applies these notions to tensors of rank 0, 1 and 2. The purpose of this application is not didactic; the cases under discussion will prove to be crucial in paragraph 4, the application to GR. The discussion of the cases is quite detailed. This could mean that Grossmann thought they were new. In any case, they cannot be found in any of the cited sources in the form presented here.

### 6.7.1 "Erweiterung" (Covariant Derivative)

Grossmann first defines the covariant derivative ("Erweiterung"). This is "the covariant (contravariant) tensor of rank $\lambda+1$, which results from a covariant (contravariant) tensor of rank $\lambda$ by 'covariant (contravariant) differentiation' ". Grossmann notes that, according to Christoffel, this is the expression

$$T_{r_1 r_2 \dots r_\lambda s} = \frac{\partial T_{r_1 r_2 \dots r_\lambda}}{\partial x_s} - \sum_k \left( \left\{ \begin{matrix} r_1 s \\ k \end{matrix} \right\} T_{k r_2 \dots r_\lambda} + \left\{ \begin{matrix} r_2 s \\ k \end{matrix} \right\} T_{r_1 k \dots r_\lambda} + \dots + \left\{ \begin{matrix} r_\lambda s \\ k \end{matrix} \right\} T_{r_1 r_2 \dots k} \right) \tag{6.4}$$

Grossmann attributes the name "covariant differentiation" to Ricci & Levi-Cività. He first defines the Christoffel symbols of the second kind after using them in equation (6.4), then those of the first kind, and then the relation between the two kinds of symbols. Finally, Grossmann introduces the "contravariant extension"; this is the operation of applying the above operation, the reciprocal, to $T$, and "raising" the differentiation index.

It is not easy to determine the source of Grossmann's notion of covariant derivative; it is useful to take the history of that concept into account. According to Luca Dell'Aglio (1996), the covariant derivative was first introduced by Christoffel in order to construct "new forms [tensors] of higher orders". However, Christoffel did not interpret the algebraically defined operation as a form of derivative, as the focus of his work was not analytic. He was merely interested in the construction of algebraic invariants. Dell'Aglio notes that it was Ricci who brought together the algebraic origin of the concept with the analytic tradition, especially the differential parameters of Lamé and Beltrami, and interpreted the expression as a generalized (coordinate independent) form of derivative.

Thus, on top of attributing the name "covariant derivative" to Ricci & Levi-Cività, Grossmann drew on Ricci's work for its interpretation. The importance of the relation to differential parameters that Ricci established will be clarified later. However, Grossmann probably did not adopt the expression of the covariant derivative, i.e. equation (6.4), from Ricci & Levi-Cività: the notation is just too different. Ricci & Levi-Cività write on p. 138:

> M. Christoffel [...] has first noted that if a system of order $m$, $X_{r_1 r_2 \dots r_m}$, is covariant, the system of order $m+1$
>
> $$X_{r_1 r_2 \dots r_m r_{m+1}} = \frac{\partial X_{r_1 r_2 \dots r_m}}{\partial x_{r_{m+1}}} - \sum_1^m l \sum_1^n q \left\{ \begin{matrix} r_l r_{m+1} \\ q \end{matrix} \right\} X_{r_1 r_2 \dots r_{l-1} q r_{l+1} r_m} \tag{6.5}$$
>
> is covariant too. We call this operation [...] *covariant derivative.*

In contrast to Grossmann, Ricci & Levi-Cività use a compact double sum in their definition. But more importantly, Ricci & Levi-Cività do not explicitly define the Christoffel symbols: they appear in the above definition without ever being defined in terms of the metric. Later, they introduce the Riemann tensor without using Christoffel's notation; instead, they devise their own notation. Thus, Grossmann must have taken the definition and properties of the Christoffel symbols from some other source.[17]

Christoffel's notation, on the other hand, is very similar to Grossmann's. He writes on p. 57:

> Under the assumption that all integrability conditions are satisfied, every complete system of transformation relations $[(\alpha_1 ... \alpha_\mu)$ $...]$ of order $\mu$ yields a new complete system of transformation relations $[(\alpha\alpha_1 ... \alpha_\mu)\ ...]$ of order $\mu + 1$ if one defines

$$(ii_1 ... i_\mu) = \frac{\partial(i_1 i_2 ... i_\mu)}{\partial x_i} - \sum_\lambda \left[ \left\{ \begin{matrix} ii_1 \\ \lambda \end{matrix} \right\} (\lambda i_2 ... i_\mu) + \left\{ \begin{matrix} ii_2 \\ i_\lambda \end{matrix} \right\} (i_1 \lambda ... i_\mu) + ... \right] \tag{6.6}$$

Thus, Grossmann's account is not based exclusively on Christoffel, nor on Ricci & Levi-Civitá. He cannot have used Bianchi exclusively either, because Bianchi does not introduce the covariant derivative as such. However, Bianchi has a very useful paragraph summarizing useful properties of the Christoffel symbols, which Grossmann might have used.

Summing up, Grossmann probably used a combination of Ricci & Levi-Civitá, Bianchi, and Christoffel for the notion of covariant derivative. He adopted the interpretation of the derivative provided by Ricci & Levi-Civitá, which is useful for generating differential operators (differential parameters). On the other hand, he adopted the notation used by Christoffel (and Bianchi), not Ricci & Levi-Civitá's.

### 6.7.2 Divergence and "Generalized Laplacean Operation"

Divergence is defined as the successive application of covariant derivative and inner product (contraction) with the fundamental tensor. This yields a new tensor of rank diminished by 1. The operation, as Grossmann notes, is not unique for general tensors, as the product can be formed with any one of the indices. The "Generalized Laplacean Operation" is the successive application of covariant derivative and divergence. The result is a tensor of the same rank.

[17] It is unclear why Ricci & Levi-Civitá use Christoffel's notation only in the context of the covariant derivative. One possibility is that this is a slip of notation. Later on, they use a more unified notation, when they introduce the Riemann tensor. In any case, Ricci & Levi-Civitá acknowledge the fundamental importance of Christoffel's work several times. The change of notation should, therefore, not be interpreted as an attempt to downplay Christoffel's importance.

What is the source of these operations? Ricci and Levi-Civita (1901, ch. 3, pp. 163.; ch. 6, pp. 191) discuss differential parameters, but not in sufficient detail; for example, no proofs are given. According to Dell'Aglio (1996), Ricci discussed differential parameters of higher orders as early as 1886. Therefore, Grossmann's results, in the discussion of the cases $n = 0, 1, 2$, are probably not original work. However, Grossmann did not use these earlier results by Ricci either. As we will see below, the most probable source, especially in the application of the operator to particular cases, is Bianchi and Lukat (1899).

### 6.7.3 Application to Scalar $T$

Grossmann discusses the notions of tensor analysis first in application to a tensor of rank 0, i.e. to a scalar $T$. He defines the gradient of $T$ and, based on this, the first Beltrami parameter. More importantly, he introduces the "generalized Laplacean operation", which is formed as the successive application of derivative and divergence. Grossmann notes that this is the second Beltrami parameter (what we would now call Laplace-Beltrami operator), and states, but does not prove, that it can be given the following form:

$$\frac{1}{\sqrt{g}}\sum_{rs}\frac{\partial}{\partial x_s}\left(\sqrt{g}\gamma_{rs}\frac{\partial T}{\partial x_r}\right) \tag{6.7}$$

Grossmann refers the reader to Bianchi for this result, and to the parallel derivation in the case of tensors of rank 1.

### 6.7.4 Application to Covariant Tensor $T_r$

Grossmann's main goal in this part is to show that the divergence of a covariant tensor (left hand side of the following equation) can be given a particular form (right hand side):

$$\sum_{rsk}\gamma_{rs}\left(\frac{\partial T_r}{\partial x_s} - \begin{Bmatrix} rs \\ k \end{Bmatrix} T_k\right) = \frac{1}{\sqrt{g}}\sum_{rs}\frac{\partial}{\partial x_s}(\sqrt{g}\gamma_{rs}T_{rs}) \tag{6.8}$$

This result is mentioned, but not derived, by Ricci & Levi-Civitá, p. 164. Grossmann mentions that Kottler (1912, p. 1679) arrived at the same result, albeit using different methods; the theory of integral forms. In any case, Grossmann finds it necessary to prove the result in the spirit of the ADC, using methods that can be found in Bianchi's book. Maybe he thought that he was the first to prove this result using the ADC methods. In any case, the result is important for Einstein's part, as we will see below.

The derivation indicates that at this point, Grossmann was already quite familiar with basic facts of tensor algebra and analysis; for example, he makes free use of the symmetry of the metric. A reconstruction of his derivation also suggests that he might have drawn on Bianchi's monograph, which summarizes some useful relations. Grossmann's derivation is very similar to Bianchi's proof that the second Beltrami differential parameter can be given a particular form.

### 6.7.5 Application to Contravariant Tensor $\Theta_{rs}$

Again, Grossmann's goal is to bring the divergence of a rank two tensor into a particular form. The starting point of the transformation is what Grossmann calls column divergence ("Kolonnendivergenz"):

$$\Theta_s = \sum_{rk} \left( \frac{\partial \Theta_{rs}}{\partial x_r} + \begin{Bmatrix} rk \\ r \end{Bmatrix} \Theta_{ks} + \begin{Bmatrix} rk \\ s \end{Bmatrix} \Theta_{rk} \right) \tag{6.9}$$

Grossmann states the following identity that is necessary for the subsequent transformation:

$$\sum_r \begin{Bmatrix} rk \\ r \end{Bmatrix} = \sum_{rs} \gamma_{rs} \begin{bmatrix} rk \\ s \end{bmatrix} = \sum_{rs} \frac{1}{2} \gamma_{rs} \frac{\partial g_{rs}}{\partial x_k} = \frac{\partial log \sqrt{g}}{\partial x_k} \tag{6.10}$$

The first identity already appears in the beginning of the second paragraph. The second identity has not been used before, and, as it is an easy result, Grossmann does not prove it. The third identity is stated in the derivation of the rank one tensor.[18] Bianchi derives and states equation (6.10) as equation 20 on p. 45. It is plausible that Grossmann took the result from there.

Putting (6.9) and (6.10) together, and recalling that $\frac{\partial log \sqrt{g}}{\partial x_k} = \frac{1}{\sqrt{g}} \frac{\partial \sqrt{g}}{\partial x_k}$ yields

$$\begin{aligned} \Theta_s &= \sum_{rk} \left( \frac{\partial \Theta_{rs}}{\partial x_r} + \frac{1}{\sqrt{g}} \frac{\partial \sqrt{g}}{\partial x_k} \Theta_{ks} + \begin{Bmatrix} rk \\ s \end{Bmatrix} \Theta_{rk} \right) \\ &= \frac{1}{\sqrt{g}} \sum_r \frac{\partial}{\partial x_r} (\sqrt{g} \Theta_{rs}) + \sum_{rk} \begin{Bmatrix} rk \\ s \end{Bmatrix} \Theta_{rk} \end{aligned} \tag{6.11}$$

This result is important for the proof, in paragraph four, that the divergence of the energy-momentum tensor is generally covariant.

Summing up, Grossmann probably considered at least some of the results in the application of the tensor analytic notions to be new, especially the cases of rank one and two. It is plausible that for much of this section, he

[18]It is unclear why Grossmann states this identity here, and not earlier, or in a more organized manner. This could indicate that Grossmann wrote the exposition in a rush.

relied on Bianchi's monograph, which contains many of the necessary tools. The same is not true for Kottler; Grossmann probably refers to him to clarify that his own result is not new, except for the derivation.

The fact that Grossmann relies on Bianchi, and not on Kottler, who uses the theory of integral forms to derive some of the results, can be explained by placing Grossmann in the methodological tradition of Ricci & Levi-Cività, who defend a "purity of method" in the introduction to their paper; see section 6.4 above.

## 6.8 Paragraph 3: Special Tensors (Vectors)

In this paragraph, Grossmann discusses antisymmetric tensors. According to Reich (1994, p. 198), this is the first definition of an antisymmetric tensor of rank $n$. Grossmann writes that "we can reduce the theory of vectors of the n-th kind (four- and six-vectors for $n = 4$) to the special tensors of rank $\lambda$. From the perspective of the general theory, it is more convenient to start with tensors and treat vectors as a special case."

Grossman introduces the Levi-Cività symbol, as he calls it, in a footnote ("Das 'System $\epsilon$' von Ricci und Levi-Cività"); $e$ the covariant, and $\epsilon$ the contravariant discriminant tensor, i.e. the Levi-Cività symbol. This is obviously taken from Ricci & Levi-Cività.

Grossmann derives two results that are important for the physical part. First, he shows that the divergence of a special tensor of rank two takes a particular form; this is a specialization of a result derived above. The divergence of a special tensor of rank two $\Theta_{\mu\nu}$ is

$$\Theta_\mu = \sum_\nu \frac{1}{\sqrt{g}} \frac{\partial}{\partial x_\nu} (\sqrt{g} \cdot \Theta_{\mu\nu}) \tag{6.12}$$

Second, he derives the "dual" of a contravariant tensor of rank two.

## 6.9 Paragraph 4: Mathematical Supplement to Part I

This paragraph has three sections. Two of these sections, the first and the third, have the simple purpose of supplementing necessary calculations for Einstein's part of the Entwurf. The second section, however, is unusual in that it does not contribute to the Entwurf theory, but discusses an "alternative" approach to the field equations. Grossmann explains, and rejects, the possibility of finding a generally covariant differential operator based on the Riemann tensor. From the modern perspective, this is the most spectacular part of the Entwurf, and also what might have led Einstein and Grossmann to qualify the paper as an "outline".

### 6.9.1 Proof of the General Covariance of the Energy-Momentum Conservation Equation

The goal of the first part is to prove the general covariance of the energy-momentum equation, which was heuristically derived in Einstein's part. This equation is, up to a factor of $\sqrt{-1}$:

$$\sum_{\mu\nu} \frac{\partial}{\partial x_\nu}\left(\sqrt{g} \cdot g_{\mu\nu} \cdot \Theta_{\mu\nu}\right) - \frac{1}{2}\sqrt{g} \cdot \sum_{\mu\nu} \frac{\partial g_{\mu\nu}}{\partial x_\sigma} \cdot \Theta_{\mu\nu} = 0 \qquad (6.13)$$

for $\sigma = 1, 2, 3, 4$.

The starting point is the result from paragraph 2, i.e. that the divergence of a contravariant tensor $\Theta_{\mu\nu}$ can be written as in equation (6.11). The ingredients of the derivation are integration by parts as well as the fact that, by symmetry of $\Theta_{\mu\nu}$, one can exchange dummy indices. The calculation is more or less straightforward. Grossmann describes the result as follows: "The divergence of the (contravariant) stress-energy-tensor of the material flow, or of the physical process, vanishes."

The result, that the energy balance equation is generally covariant, is new. The main tool is the fact that the divergence of a contravariant tensor can be written as in equation (6.11). We argued there that Grossmann probably relied on Bianchi for this result.

### 6.9.2 Differential Tensors of a Manifold Given by its Line Element

In this interesting section, Grossmann explores, and rejects, the mathematical approach to the field equations, which will later prove to be correct. We have identified three central aspects of this passage. We will compare Grossmann's account of these three points with the corresponding accounts in the works of Ricci & Levi-Cività, Christoffel, Bianch, and Riemann.

1. *Riemann Tensor Formula:* Grossmann defines the Riemann tensor in two forms: i) in fully covariant form, and ii) with one contravariant index. He notes the relation between the two expressions. The Riemann tensor is one of the fundamental objects in the construction of the later, correct, field equations. We will compare the possible sources of Grossmann's account.

2. *Interpretation of the Riemann Tensor:* Grossmann notes that, based on the Riemann tensor (and the discriminant tensor), we can construct the complete system of differential tensors of the line element, with generally covariant algebraic and differential operations. Grossmann emphasizes the "eminent importance of these concepts for differential

geometry". This characterization of the Riemann tensor is traced back to the mathematical literature.

3. *(Failed) Application of the Riemann tensor in GR:* Grossmann constructs what we now recognize as the Ricci tensor in equation (6.23), but ultimately rejects it as an operator for the field equations, because he is convinced that it fails to reduce to the Newtonian limit in a suitable manner. These considerations do not go back to the mathematical literature.

**Grossmann's Program**

In the beginning of the second section, Grossmann states his program:

> The formulation of the differential equations of the gravitational field [...] directs our attention to the differential invariants and differential covariants of the quadratic differential form [$g_{\mu\nu}$]. The theory of these differential invariants [...] leads to the differential tensors given by a gravitational field. The complete system of these differential tensors for arbitrary transformations can be reduced to a differential tensor of rank four found by Riemann [fn.: Riemann, Ges. Werke, S. 270] and, independently, by Christoffel [fn: Christoffel, 1.c., S. 54]; we will call it the Riemann differential tensor [...]. By covariant algebraic and differential operations, the Riemann differential tensor and the discriminant tensor [..] provide the complete system of differential tensors (and thereby also of differential invariants) of the manifold.

Grossmann establishes a direct line of reasoning from the differential equations of a gravitational field, via the differential tensors based on the metric, to the Riemann tensor, which allows the construction of all differential tensors of this kind. If you want to find generally covariant differential equations based on the metric, the Riemann tensor is *the* mathematical object you need. We will return to the stated interpretation of the Riemann tensor below. Note the (faulty) reference to Riemann, and the (correct) reference to Christoffel.

**Grossmann: Riemann Tensor Formula**

Grossmann writes the fully covariant Riemann tensor as follows:

$$\begin{aligned} R_{iklm} = (ik, lm) = \frac{1}{2}\left(\frac{\partial^2 g_{im}}{\partial x_k \partial x_l} + \frac{\partial^2 g_{kl}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{il}}{\partial x_k \partial x_m} - \frac{\partial^2 g_{mk}}{\partial x_l \partial x_i}\right) \\ + \sum_{\rho\sigma} \gamma_{\rho\sigma}\left(\begin{bmatrix} im \\ \rho \end{bmatrix}\begin{bmatrix} kl \\ \sigma \end{bmatrix} - \begin{bmatrix} il \\ \rho \end{bmatrix}\begin{bmatrix} km \\ \sigma \end{bmatrix}\right) \end{aligned} \tag{6.14}$$

Grossmann notes that $(ik, lm)$ is also called the Christoffel four-index symbol of the first kind. Then he introduces the four-index symbol of the second kind:

$$\{ik, lm\} = \frac{\partial \left\{ {il \atop k} \right\}}{\partial x_m} - \frac{\partial \left\{ {im \atop k} \right\}}{\partial x_l} + \sum_\rho \left( \left\{ {il \atop \rho} \right\} \left\{ {\rho m \atop k} \right\} - \left\{ {im \atop \rho} \right\} \left\{ {\rho l \atop k} \right\} \right) \quad (6.15)$$

Grossmann notes that the symbol of the second kind, a different form of the Riemann tensor, is a mixed tensor of covariant rank three and contravariant rank one. The two kinds of four-index symbols are related as follows:

$$\{ik, lm\} = \sum_k \gamma_{\rho k} (ik, lm) \quad (6.16)$$

$$(ik, lm) = \sum_\rho g_{k\rho} \{i\rho, lm\} \quad (6.17)$$

**Riemann: Riemann Tensor Formula**

In Riemann's "Commentatio" (Riemann, 1876b, p. 381, eq. I), we can find the following expression:

$$\begin{aligned} &\frac{\partial^2 b_{\iota,\iota''}}{\partial s_{\iota'} \partial s_{\iota'''}} + \frac{\partial^2 b_{\iota',\iota'''}}{\partial s_\iota \partial s_{\iota''}} - \frac{\partial^2 b_{\iota,\iota'''}}{\partial s_{\iota'} \partial s_{\iota''}} - \frac{\partial^2 b_{\iota',\iota''}}{\partial s_\iota \partial s_{\iota'''}} \\ &+ \frac{1}{2} \sum_{v,v'} \left( p_{v,\iota',\iota'''} p_{v',\iota,\iota''} - p_{v,\iota,\iota'''} p_{v',\iota',\iota''} \right) \frac{\beta_{v,v'}}{B} = 0 \end{aligned} \quad (6.18)$$

Riemann writes the components of the expression on the l.h.s. as $(\iota\iota', \iota''\iota''')$. This is the Riemann tensor. Its vanishing, i.e. equation (6.18), is a necessary condition for a flat metric, i.e. we can transform it into $\sum_\iota dx_\iota^2$.[19]

Riemann's notation is quite different from Grossmann's, especially in comparison with the other sources. Riemann does not use the inverse metric, and he does not use Christoffel's notation either.[20] Furthermore, the approach taken in the "Commentatio" is not relevant for Grossmann – for example, Riemann is not interested in the tensorial nature of the Riemann tensor. This further supports our argument in section 6.3 that Riemann was not directly consulted, or used by, Grossmann.

[19] See also Zund (1983, p. 87).

[20] According to Zund (1983), he could not have known Christoffel's work, as Riemann wrote the "Commentatio" around 1861, before Christoffel's 1869 paper. However, Christoffel did not know the "Commentatio" either, as it only appeared in print in the 1876 edition of Riemann's works. The two are independent.

### Ricci & Levi-Cività: Riemann Tensor Formula

Ricci & Levi-Cività discuss the Riemann tensor, or "Riemann system" ("système de Riemann"), as they call it, twice. On pp. 142., they define the Riemann tensor in two forms and state some important properties. To give an example, they note that the Riemann tensor is, what we would now call, the commutator of the covariant derivative. This passage contains the first citation of Riemann.[21] They also cite Christoffel, but without a page number. Grossmann, on the other hand, gives page numbers in his Christoffel citations. This is a further clue that Grossmann did consult Christoffel separately. On pp. 160, Ricci & Levi-Cività discuss the construction of differential invariants from the Riemann tensor.

Ricci & Levi-Cività did consult Riemann's "Commentatio". However, it is highly unlikely that Grossmann has taken his account of the Riemann tensor from Ricci & Levi-Cività. There are two reasons for this claim. Firstly, the notation used by Ricci & Levi-Cività, on the one hand, and Grossmann on the other, do not match – most strikingly, as I mentioned above, they only use Christoffel symbols once in the whole paper, in the definition of the covariant derivative. Later, they even introduce their own notation for the Christoffel symbols (p. 142):

$$2a_{rs,t} = \frac{\partial a_{rt}}{\partial x_s} + \frac{\partial a_{st}}{\partial x_r} - \frac{\partial a_{rs}}{\partial x_t} \tag{6.19}$$

$a_{rs}$ are the components of the metric. Subsequently, they write the fully covariant Riemann tensor ("système covariant de riemann") as follows:

$$a_{rs,tu} = \frac{\partial a_{rt,s}}{\partial x_u} - \frac{\partial a_{ru,s}}{\partial x_t} + \sum_{pq=1}^{n} a^{(pq)}\left(a_{ru,p}a_{st,q} - a_{rt,p}a_{su,q}\right) \tag{6.20}$$

Secondly, Ricci & Levi-Cività do not discuss the four-index symbol of the second kind at all, i.e. they do not introduce the (general) Riemann tensor with one contravariant index, and, consequently, the relations between the two expressions of the Riemann tensor are not stated as well. They only discuss a contravariant version of the Riemann tensor in the case $n = 3$.

### Christoffel: Riemann Tensor Formula

Grossmann cites Christoffel's paper as one of the origins of the Riemann tensor, with a specific reference to p. 54. Grossmann's Riemann tensor is more or less the same as Christoffel's. Christoffel writes the fully covariant Riemann tensor (p. 54, eq. 14) as

[21] This is the faulty reference to p. 270 instead of p. 370; see our discussion in section 6.3 above.

$$(gkhi) = \frac{1}{2}\left(\frac{\partial^2 w_{gi}}{\partial x_h \partial x_k} + \frac{\partial^2 w_{hk}}{\partial x_g \partial x_i} - \frac{\partial^2 w_{gh}}{\partial x_i \partial x_k} - \frac{\partial^2 w_{ik}}{\partial x_g \partial x_h}\right) + \sum_{\alpha\beta} \frac{E_{\alpha\beta}}{E}\left(\begin{bmatrix} gi \\ \alpha \end{bmatrix}\begin{bmatrix} hk \\ \beta \end{bmatrix} - \begin{bmatrix} gh \\ \alpha \end{bmatrix}\begin{bmatrix} ik \\ \beta \end{bmatrix}\right) \tag{6.21}$$

This expression deviates from Grossmann's in the order of some indices, Christoffel's notation for the metric ($w_{\alpha\beta}$), the use of the subdeterminant divided by the determinant ($\frac{E_{\alpha\beta}}{E}$) instead of the inverse metric, and the missing comma in the four-index symbol. It is probable that Grossmann's account is based directly on Christoffel, or on a source close to Christoffel.

However, Christoffel is probably not the only source. Most importantly, Christoffel does not define the four-index symbol of the second kind – it is, strictly speaking, not a Christoffel symbol. Christoffel writes the Riemann tensor with one contravariant and three covariant indices on p. 52, but he does not introduce a special notation for it; instead he "lowers" the contravariant index in order to get the fully covariant Riemann tensor. Christoffel also does not state the transformation relations between the two forms of the Riemann tensor, (6.16) and (6.17) above.

**Bianchi: Riemann Tensor Formula**

Bianchi writes the four-index Christoffel symbol of the first kind as follows (p. 51, eq. 32*)

$$(rk, ih) = \frac{1}{2}\left(\frac{\partial^2 a_{rh}}{\partial x_i \partial x_k} + \frac{\partial^2 a_{ik}}{\partial x_r \partial x_h} - \frac{\partial^2 a_{ri}}{\partial x_h \partial x_k} - \frac{\partial^2 a_{hk}}{\partial x_r \partial x_i}\right) + \sum_{lm} A_{lm}\left(\begin{bmatrix} rh \\ m \end{bmatrix}\begin{bmatrix} ik \\ l \end{bmatrix} - \begin{bmatrix} ri \\ m \end{bmatrix}\begin{bmatrix} hk \\ l \end{bmatrix}\right) \tag{6.22}$$

This is very similar to Grossmann's notation, except for the metric ($a_{ik}$) and inverse metric ($A_{ik}$) in place of Grossmann's $g_{\mu\nu}$ and $\gamma_{\mu\nu}$. Bianchi and Grossmann both use the comma in the middle of the four-index symbol. The motivation for the introduction of the comma is probably to indicate its symmetries.

Bianchi is the only one among Grossmann's sources to introduce the four-index symbol of the second kind (p. 49, eq. 27). What is more, he also states the relation between the two kinds of four-index symbols (p. 49, eqs. 29, 29*), which is very similar to Grossmann's equations (6.16) and (6.17) above.

Bianchi's presentation is closest to Grossmann's; it is almost a perfect match. It is very likely that Grossmann used Bianchi, and probably also

Christoffel. However, a close parallel reading of Bianchi and Christoffel reveals that Bianchi took big parts of his account directly from Christoffel: Much of paragraphs 27 (on the four-index symbols) and 28 (on the curvature of a binary differential form) are parallel to Christoffel's parts 4 and 5. In his paragraphs 27 and 28, Bianchi reconstructs Christoffel in a more readable way. In sum, at least here, Bianchi is a mediator between Christoffel and Grossmann.[22]

Riemann, on the other hand, was not as big an influence on Bianchi as Christoffel. Bianchi does not list Riemann as a major influence on chapter two, and while Riemann's habilitation paper is in Bianchi's bibliography, Bianchi does not mention the "Commentatio" at all.[23]

### Grossmann: Interpretation of the Riemann Tensor

Grossmann's interpretation of the Riemann tensor has two aspects. First, the Riemann tensor has a systematic mathematical significance. It produces *all* differential tensors and invariants of a "manifold given by its line element". In modern terms, it delivers all generally covariant differential operators that can enter into the field equations. This is the main mathematical message.[24]

Second, he points out the "eminent importance" of the Riemann tensor for differential geometry, and mentions one particular property of the Riemann tensor in a footnote: "[t]he identical vanishing of the tensor $R_{iklm}$ constitutes the necessary and sufficient condition for the transformability of the differential form into $\sum_i dx_i^2$." For Grossmann, this seems to establish the connection between the Riemann tensor, a purely mathematical object, and the problem of finding differential equations of the gravitational field: the vanishing of the Riemann tensor is equivalent to the possibility of transforming the metric into the form $\sum_i dx_i^2$, in which case we get a flat metric.

We will now try to locate these aspects in the mathematical literature.

### Ricci & Levi-Cività: Interpretation of the Riemann Tensor

Ricci & Levi-Cività discuss the construction of differential invariants based on the Riemann tensor in chapter three, "On analytical applications", pp. 160, par. 2. Their objective is the following:

[22]This does in no way mean that Bianchi plagiarized Christoffel; the latter's influence is acknowledged: Bianchi cites Christoffel's paper as one of the main sources used for chapter two.

[23]The bibliography reads: "Riemann, Über die Hypothesen, welche der Geometrie zu Grunde liegen. (Gesammelte Werke, Leipzig -Teubner, S. 254. Vgl. auch die Zusätze von Dedekind, S. 517)". The "Commentatio" is not mentioned. He draws attention to remarks by Dedekind that were added later. The remarks ("Zusätze") by Dedekind on p. 517 are part of a biographical sketch and concern the genesis of Riemann's habilitation paper.

[24]This result was first proved in Christoffel (1869). As stated, the result is not entirely correct and has to be hedged suitably; see Ehlers (1981) for an appraisal of Christoffel's result and later generalizations.

> Given a definite quadratic form $\phi$ [i.e. the fundamental tensor] and any number of associated systems $S$ (covariant or contravariant) [i.e. tensors], determine all the absolute invariants that can be formed with the coefficients of $\phi$, the elements of the systems $S$, and their derivatives up to a fixed order $\mu$.

Ricci and Levi-Civita do not define absolute invariants. According to Reich (1994, p. 55), the definition of absolute invariants is due to Siegfried Aronhold: It is a transformation that is independent of the substitution determinant. In modern terminology, an absolute invariant transforms like a tensor, while invariants in general transform as tensor densities. In this respect, Ricci & Levi-Cività deviate from Grossmann, who includes the use of the discriminant tensor in the construction of differential invariants. Ricci & Levi-Cività then state the answer to this problem:

> To obtain all the absolute differential invariants of order $\mu$, it is sufficient to determine the algebraic invariants of the systems of the following form:
>
> 1. fundamental form $\phi$;
> 2. associated forms $S$ and their derivatives with respect to $\phi$ up to order $\mu$
> 3. (for $\mu > 1$) quadrilinear form, with coefficients from the Riemann system [i.e. Riemann tensor]; derivatives of the form up to order $\mu - 2$
>
> *Proper invariants* are those invariants of a form $\phi$ that depend on the form $\phi$ and its derivatives only; we can deduce the following two corollaries from the preceding proposition:
>
> - The forms of class 0 do not admit of any proper differential invariants
> - The forms of higher classes do not have differential invariants of order one; their invariants of order $\mu > 1$ are those of the forms 1) and 3).

They then go on to discuss the cases $n = 2, 3$.[25]

Ricci & Levi-Cività's discussion of invariants generated by the Riemann tensor is similar to Grossmann's in that they discuss how the *complete* set of (absolute) differential invariants can be obtained. The terminology, however, differs from Grossmann's in several respects. First, they state the theorem in terms of absolute invariants, a notion that Grossmann does not use. The notion of algebraic invariant is not properly defined as well. We can guess that

[25] Curiously, they announce that they want to come back to these issues in the chapter on geometric applications, specifically in par. 8; however, this paragraph does not exist.

they allude to the tensor algebraic operations defined in the first chapter. We can deduce, from the second corollary, that those differential invariants of order $\mu = 2$ that can be constructed from the metric alone are the algebraic invariants constructed from the metric and the Riemann tensor; there is no need to form derivatives of the latter. Secondly, and perhaps more importantly, Ricci & Levi-Cività do not discuss constructions involving tensor densities.[26]

Ricci & Levi-Cività mention on p. 143 that if we can transform the fundamental tensor into the form $\sum_i dx_i^2$, the Riemann tensor vanishes identically; this is repeated on p. 161. They do not state that it is an equivalence. As pointed out above, they note that the Riemann tensor is basically the commutator of the covariant derivative.

In sum, it is implausible that Grossmann based his account exclusively on Ricci & Levi-Cività.

### Christoffel: Interpretation of the Riemann Tensor

In contrast to Ricci & Levi-Cività, Christoffel treats the case of invariants with nontrivial substitution determinant, i.e. tensor densities:

> [...] the system containing all the necessary and sufficient conditions for the transformation from $F$ to $F'$ is $I' = r^\lambda I, I'_1 = r^{\lambda_1} I_1, ...$; with $r$ substitution determinant, and $\lambda$ constant. It is appropriate to call these invariants of the forms $F, G_4, G_5, ...$ [i.e. the fundamental form, the Riemann tensor, and its covariant derivatives] the complete system of invariants of the differential expression $F$.

It is plausible that Grossmann consulted the Christoffel paper directly on this point.

### Bianchi: Interpretation of the Riemann Tensor

Bianchi discusses differential invariants and covariants in the following passages: In paragraph 22, he defines the differential invariants and differential parameters of a quadratic differential form. Paragraph 24 is entitled "Equivalence of two quadratic differential forms". Here he derives what he calls the fundamental equation (equation I on p. 43). This equation expresses the second differential quotients of the coordinate system $x$ in terms of the first

[26] Ricci & Levi-Cività's discussion is not really user-friendly. The above theorem is stated without much explanation or proof; they only refer to earlier papers by Ricci. Concepts such as absolute and algebraic invariants are not introduced. As a consequence, it is hard to understand how exactly we can construct the proper differential invariants we are after. We assume that Ricci & Levi-Cività are referring to proper differential invariants in the second corollary as well.

differential quotients. It can be found one-to-one in Christoffel, p. 49, equation (9). In paragraph 27 on "four-index symbols", i.e. the Riemann tensor, he discusses the construction of differential invariants, as opposed to differential parameters. He writes: "We now want to build a covariant of degree four in the differentials, whose coefficients are built from the fundamental form and of its (first and second) derivative. In this manner we are able to construct differential invariants." However, he does not state, or discuss, whether the differential invariants we get in this manner form a complete system of such invariants.

In sum, Bianchi does not prove statements about the construction of differential invariants in chapter two, at least not in the generality required by Grossmann.[27]

### Ricci Tensor: The (Failed) Application

The last part of Grossmann's discussion of a generally covariant approach to gravity is the one that caused the most controversy and has been widely discussed by historians of general relativity.[28] The mistake we find documented in this short passage sent Einstein on a more than three-year-long odyssey until he finally found the correct field equations. Here is the notorious passage in full:

> Indeed, it is possible to find a covariant differential tensor of rank two and order two, $G_{im}$, that could enter into these equations, namely
>
> $$G_{im} = \sum_{kl} \gamma_{kl}(ik, lm) = \sum_{k} \{ik, km\} \tag{6.23}$$
>
> However, it turns out that in the special case of an infinitely weak static field, this tensor does not reduce to the expression $\Delta\phi$. Therefore, we have to leave open how far the general theory of differential tensors linked to the gravitational field is related to the problem of the gravitational equations. Such a relation would have to exist, if the gravitational equations allowed for arbitrary substitutions; however, in this case it seems impossible to find differential equations of order two. In contrast, if it turned out that the gravitational equations only allow a certain group of transformations, it would be reasonable if the general differential tensors will not do. As stated in the physical part, we are unable to give our opinion on these matters.

[27] The annotations in Einstein and Grossmann (1995) refer us to Bianchi for the passages on complete systems of differential invariants, see annotations [69], [70]. In view of the above, these references are misleading.

[28] See Janssen et al. (2007a,b) for references.

In the first sentence, Grossmann states two formal requirements on the differential operator that could enter into the field equation. The requirement that the object should be of rank two is suggested by the right hand side of the field equations: the EM tensor is of rank two; this should be mirrored on the left hand side. The restriction of the differentials to order two is discussed in Einstein's part. It is justified by the analogy to the Poisson equation.[29]

The requirement that the operator be of rank two excludes the Riemann tensor as a candidate, but, as Grossmann noted above, other candidate operators can be generated from the Riemann tensor. In equation (6.23) he therefore contracts two indices of the Riemann tensor, thus reducing the number of free indices to two. This results in, what we now know as, the Ricci tensor.

It is possible that Grossmann found the Ricci tensor on his own: the Ricci tensor has not been noted in any of the writings that Grossmann used.[30] If this is true, it is quite an achievement. However, it should also be noted that the Ricci tensor is not the only tensor that satisfies the above requirements; there is the possibility of further algebraic operations that lead, in the end, to the correct field equations.[31] This possibility is nowhere noted. Grossmann had probably not yet achieved a sufficient understanding of tensor calculus to see this alternative.

In the next sentence, Grossmann states the reason why he and Einstein rejected the Ricci tensor as a suitable differential operator: it fails to yield the correct expression $\Delta\phi$, i.e. the Laplace operator entering into the Poisson equation, in the weak, static limit. Grossmann's very brief account is, of course, not sufficient to understand or reconstruct the reasoning that ultimately led to the rejection of the Ricci tensor, and to Einstein's year-long rejection of generally covariant field equations. However, it is possible to reconstruct Einstein's (and Grossmann's) reasons for rejecting the Ricci tensor at this point by taking other sources into account, in particular the famous Zurich notebook.[32]

[29]Einstein seems not to be entirely satisfied with this justification; he admits that this is unsatisfactory and that there is no *a priori* reason to exclude differential operators of higher degrees.

[30]The Ricci tensor was first discussed in Ricci (1904). However, Grossmann was probably not aware of this paper.

[31]The differential opterator entering into the Einstein field equations is $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$, where $R$, the Ricci scalar, can be obtained from the Ricci tensor, by contracting the two indices. The cosmological term was probably neglected by Einstein and Grossmann, because the Poisson equation does not suggest such a term; by analogy, the generally covariant candidate does not contain it either.

[32]The first account of "Einstein's Odyssey" is in Norton (1984); the story is thoroughly discussed in Janssen et al. (2007a,b). We will not further explore this fascinating part of the genesis of general relativity in this chapter, as it is only indirectly tied to Grossmann's mathematical sources. A short account of the story can be found in chapter 5; a philosophical discussion in chapter 8.

In the rest of the section, Grossmann expresses ambivalence about the prospect of a generally covariant formulation of gravitational theory. He seems to think that this is not a mathematical question, but has to be settled based on physical considerations.

#### Differential Tensors of a Manifold Given by its Line Element: Summary

We are now in a position to assess the relative importance of Grossmann's mathematical sources and their influence on 1) the expression of the Riemann tensor, 2) the interpretation of the Riemann tensor, and 3) the construction of the Ricci tensor and its rejection as a candidate differential operator for the field equatinon

First, the origin of Grossmann's Riemann tensor is, in all probability, Bianchi's textbook. Bianchi, in turn, based his account on Christoffel. We thus have a direct line of influence from Christoffel to Grossmann. On the other hand, the same can probably be excluded for Ricci & Levi-Civitá, and Riemann. In all probability, Grossmann did consult Ricci & Levi-Civitá, and use them as a guide to the literature, at least to copy the faulty reference to Riemann.

Second, the interpretation of the Riemann tensor that allows the construction of all differential operators related to the metric tensor: Here the origin is not as clear cut. In all probability, Ricci & Levi-Civitá served as guides, but they do not give the full story. Christoffel is the most probable source, as he is the only one discussing tensor densities in this context. Bianchi does not discuss the interpretation of the Riemann tensor in sufficient detail.

Third, we see no precedent for the Ricci tensor and the failed transition to the classical limit in the mathematical literature – this is not very surprising. However, there is a 'gap" in the mathematical literature at this point. All mathematical sources formulate the possibility of generating new candidate differential operators at an abstract level – it is not trivial to operationalize this step. As a consequence, Grossmann does not exhaust the mathematical possibilities for constructing candidate differential operators, and in particular, he does not consider the Einstein tensor as a live option. This would have been possible, had the mathematical literature been more user-friendly.

### 6.9.3 On the Derivation of the Gravitation Equations

In the third section of the "mathematical supplement", Grossmann provides calculations for the derivation of the (faulty) Entwurf field equations in Einstein's part. The steps of the derivation are not difficult: all that is needed is integration by parts, the derivative of the square root of the determinant, and rewriting the partial derivative of the inverse metric in terms of the derivative of the metric, a result that Grossmann proved in part two. The motivation for this calculation, and its relevance, have to be found in the physical part. Grossmann's part of the Entwurf ends here.

## 6.10 Grossmann and the Mathematicians: Main Lessons

In the introduction, we asked two sets of questions. The first set of questions concerns Grossmann's "passive" contributions: his mathematical sources, the mathematical traditions behind the sources, and the relative importance of the sources, theories, and traditions. The second set of questions is about Grossmann's "active" contributions. What did Grossmann add to the already existing mathematics? What are his amendments, notational and conceptual innovations, or even new results? Finally, we are also interested in the problems of the application process: Were there any misunderstandings, blunders, or even mistakes, either due to the existing mathematical literature, or to Grossmann, or even both? In this concluding section, we will answer these questions.

### 6.10.1 "Passive" Contributions: Sources, Traditions, and their Relative Importance

We examined the influence of five mathematical sources on Grossmann: Riemann, Christoffel, Ricci & Levi-Civitá, Kottler, and Bianchi. We are now in a position to gauge to what degree they influenced Grossmann's work.

First, we argued that there is no direct influence by **Riemann**, as Grossmann may not have consulted Riemann's works at all; see section 6.3 above. The same may be true for Riemann's manifold concept; see section 6.5.

Second, we did not find many traces of **Kottler's** work in the Entwurf. Grossmann emphasizes that his approach to differential parameters is different from Kottler's. The reason why Kottler is cited may be due to a question of priority; see section 6.7.4.

We found more traces of the other three sources, Christoffel, Ricci & Levi-Civitá, and Bianchi. We examined which of these Grossman used in each part of the Entwurf. Here is a summary.

**Christoffel's** contribution stands out as particularly important. He contributed crucial technical innovations that had a major impact, not only on Grossmann, but on all later developments in tensor calculus and differential geometry. We argued that Grossmann might have taken the expression for the covariant derivative directly from Christoffel; see section 6.7.1. In addition, Christoffel is probably the source of the (invariant-theoretic) interpretation of the Riemann tensor; see section 6.9.2. Christoffel's paper is not an easy read. Almost every section contains important new results, but it is not easy to figure out what exactly is going on. It is therefore no wonder that other mathematicians took up his ideas and presented them in a more accessible manner.

**Ricci & Levi-Cività's** paper has several important roles. First, it has survey character and might have served as an entry point to the mathematical literature – in this sense, Ricci & Levi-Cività's contribution is a success story. Second, they fleshed out Christoffel's work and developed it into a calculus, the ADC. They emphasize the importance of a clear, accessible presentation and notation. Third, the paper is the source of some crucial concepts of the Entwurf: it is probably the origin of Grossmann's concept of manifold, see section 6.5, tensors (the concept, not the name), see section 6.6.2, and the covariant derivative (which is based on an expression found by Christoffel), see section 6.7.1. Fourth, their paper also contains quite some notational innovations, but these were almost all dismissed or ignored by Grossmann – this would only change later, when their use of upper and lower indices, to distinguish contravariant and covariant components, became an industrial standard.

**Bianchi's** textbook on differential geometry served two main purposes. First, Grossmann almost invariably used Bianchi's notation: for example variable differentials and the inverse metric; see section 6.6.1, and, more importantly, Bianch's formula for the Riemann tensor, see section 6.9.2. Bianchi covers much of Christoffel's paper in a well-organized "textbook" style, and is much more accessible than Christoffel. Second, Bianchi provided Grossmann with a template for the construction of generalizations of Laplace-Beltrami operators; see section 6.7.3 and the following. These are indispensable tools for the crucial paragraph four, the mathematical supplements. Bianchi's textbook is not directly linked to Ricci & Levi-Cività's paper. We can speculate that Grossmann was familiar with the book before he searched the mathematical literature for Einstein, as Bianchi covers non-Euclidean geometries with constant curvature, a topic Grossmann was interested in. It should be noted that, while Bianch's book is about differential geometry, Grossmann only drew on the second chapter, which is largely on quadratic differential forms, and almost free of geometrical notions.

We have now identified the three central sources of the pure mathematics applied in Grossmann's part of the Entwurf. All three contributions point to one source as the origin, and this is Christoffel's paper. We propose that Christoffel, and part of Ricci & Levi-Cività's paper, as well as Bianchi's second chapter, can be grouped into a single tradition, an algebraic, algorithmic tradition, which focuses on the solution of a particular technical problem without immediate interpretation – Christoffel's solution of the equivalence problem with algebraic invariant theory – and on the formulation of a calculus with no particular application in mind – Ricci & Levi-Cività's ADC.

This algebraic-algorithmic tradition should be contrasted with a geometric tradition that has its origin in Gauss's work, and was extended by Riemann. We can find traces of this tradition in virtually all of Grossmann's sources. Bianch-Lukat is a textbook on differential geometry; Ricci & Levi-Cività's paper has sections on the application of the ADC to intrinsic geometry. Even Christoffel acknowledges, albeit very briefly, that the problem he is interested in has its origin in surface theory. However, we think that, for better or worse, this geometric tradition left almost no traces in the Entwurf.

Grossmann made very selective use of his sources: While he drew heavily on both Ricci & Levi-Cività and Bianchi-Lukat, the citations indicate that he did not use the geometric parts of their work; the references are all to the "algorithmic" parts and chapters. This seems plausible, in view of the fact that Grossmann's goal was to do without geometric concepts, as he writes in the introduction.

Geometric reasoning and geometric interpretations of most algebraic expressions was almost entirely neglected. The single most important object, the line element, has a natural geometric interpretation, but other than that, the geometric significance of virtually everything else is unspecified, and left unexplained, in the mathematical part of the Entwurf.

In sum, we can discern an algorithmic-algebraic tradition, starting with Christoffel, and fleshed out by Ricci & Levi-Cività, on the one hand, and Bianchi on the other, as the single most important mathematical tradition featuring in the first application of tensor calculus in the Entwurf, and also an almost complete lack of geometric interpretations of the parts of the algorithm.

### 6.10.2 "Active" Contributions

Let us now turn to the second set of questions, Grossmann's own contributions to the mathematical part of the Entwurf. We have grouped Grossmann's innovations into several categories: a) changes and innovations on a conceptual level, b) changes and innovations of notation, c) genuinely new mathematical results, and d) missed opportunities and mistakes.

**Reinterpretation of Concepts:** Implicitly, the mathematical part is the first account of the mathematics of a space-time manifold. However, Grossmann does not really spell this out; space-time interpretations of the mathematics are missing, except for the line element. We argued in section 6.5 that Grossmann might have had a quite rudimentary manifold concept; at times, he identifies it with what is given by the line element.

Grossmann introduced an extended concept of tensors, transcending its prior, physically-grounded interpretation by relabelling the covariant and contravariant "systems" of the ADC as tensors; see section 6.6.2.

Arguably the most important "interpretational" contribution is Grossmann's realization that the Riemann tensor is the mathematical object that provides candidate, generally covariant, differential operators that could enter into the field equations. The requirements on this object had been provided by Einstein, and were heuristically motivated, but were also sufficient to identify the Riemann tensor.

**New Concepts:** Grossmann introduced several new concepts, especially generalizations of existing mathematical concepts. Examples are general mixed tensors, see section 6.6.2, general antisymmetric tensors, see section 6.8, and probably the notion of general differential operators, see section 6.7.2.

**Notational Changes and Innovations:** Grossmann introduced new notation for covariant, contravariant, and mixed tensors, by assigning them three different kinds of letters; see section 6.6.2. This notation did not survive; it is more awkward than Ricci & Levi-Cività's index position system, and we cannot conclusively answer why he did not adopt their convention.

**New Mathematical Results:** A first set of new results can be found in paragraph 2, the application of tensor calculus to tensors of rank 0-2. Grossmann derives special forms of generally differential operators, see sections 6.7.4 and 6.7.5. The operator for tensors $T_r$ was known before, but Grossmann derives the result in an alternative, potentially new manner. The operator for tensors $\Theta_{rs}$ is probably also new. Both results are necessary for the results in paragraph four. Furthermore, Grossmann specializes these results to "special tensors" (antisymmetric tensors); see section 6.8.

The most important innovations can be found in the "Mathematical Supplement". First, there is the proof of the general covariance of the Energy-Momentum Conservation equation; see section 6.9.1. This is one of the contact points of physics and mathematics: a physically motivated equation is proven to be tensorial. Second, there is the search for an operator for the field equation based on the Riemann tensor. Grossmann derives the Ricci tensor as a suitable candidate operator based on algebraic operations; see section 6.9.2. Here he draws on the mathematical literature that provides general recipes of how new operators can be derived. Third, Grossmann derives the Entwurf field equations. This derivation is not difficult on a technical level, but indispensable for the physical part.

**Unfinished Business, Entwurf Character:** Some of Grossmann's formulations suggest that he was not entirely happy with his approach and his results. One example is the notation for covariant, contravariant, and mixed tensors: He emphasizes that Ricci & Levi-Cività's solution has many advantages, but that he was forced to choose a different path; see section 6.6.2.

The clearest example is Grossmann's dissatisfaction with the failed generally covariant approach to the operator entering into the field equation; see section 6.9.2. This whole section, which discusses an alternative account that has not been carried out successfully, would be superfluous if Grossmann and Einstein really trusted their solution, the Entwurf operator. The fact that this section has been included nevertheless goes a long way towards explaining why Einstein and Grossmann considered this to be an Entwurf theory.

**Missed Opportunities, Mistakes:** There are not many mistakes in Grossmann's part of the Entwurf. We could not find any mistake in the calculations that are actually carried out. The presentation of the material, however, is not very accessible and balanced. For example, the part on differential operators; see section 6.7.3, is somewhat awkward in that at times, Grossmann states auxiliary results after he has already used them. We have already pointed out that there are some missed opportunities when it comes to notation, especially the step back from Ricci & Levi-Cività's index position notation.

Finally, there is the missed opportunity with the failed application of the generally covariant approach based on the Riemann and the Ricci tensor, see section 6.9.2. Grossmann did not exhaust all candidate differential operators that can be constructed according to the mathematical literature; in particular, he (probably) missed the Einstein tensor, the correct differential operator that enters into the field equations. On the other hand, the mathematical literature was not very user-friendly in providing the users of the ADC more hands-on recipes for the construction of differential operators based on the Riemann tensor. In the end, the mistake that led to the abandonment of the generally covariant approach cannot be attributed to Grossmann or the mathematicians; the story is more complicated.

### 6.10.3 Systematic Take-Home Message

We will not flesh out the philosophical or systematic significance of our historical results here; rather, we will use the present historical study to refine an existing account of the application of mathematics in chapter 8. Nevertheless, we would like to point out some features of this case of application of mathematics that are particularly noteworthy.

First, at least in the present case, the application of mathematics is not "plug-and-play". Mathematical theories are not sitting in a shelf, waiting to be put to work. Even though the theory in question had been formulated for some time, it still took considerable effort on Grossmann's part to adapt and extend the mathematics for its application to gravitation. Not only did he have to consult multiple sources, but he also had to contribute substantive mathematical results himself. This is a first sense in which mathematics is not static.

Second, there is a contribution of mathematics to its own applicability: Mathematicians put a considerable effort into fleshing out technical work. Ricci & Levi-Cività, as well as Bianchi-Lukat, are examples. The mathematicians also tried to find applications of their work both in pure mathematics and in application. At least some mathematicians do not conceive of their contributions as free from practical considerations. This is a second sense in which mathematics is not static.

Third, there are failures and missed opportunities, both on the side of pure mathematics and its application. Even though some of the mathematicians made an effort to present their work in an accessible manner, and anticipated some of the challenges of application, there are also gaps in their presentation; at times, it is very hard to figure out what is going on even in Ricci & Levi-Cività's paper, which is supposed to be a survey paper and thus particularly accessible.

Fourth, there are systematic lessons from the particular mathematical theory, and tradition, that we found to be at the heart of the first application of the ADC to GR. At first, it could be surprising that geometry is not the most important mathematical tool in the first application of generally covariant methods to GR, and that the abstract, algebraic-algorithmic tradition, originating with Christoffel, had center stage. We think, however, that there is a good explanation for this choice; on the other hand, it also led to difficulties that plagued the early history of GR.

An obvious advantage of the algorithmic approach was that it was available as an uninterpreted calculus, rather than as a calculus of geometry. It would have taken too much time to work out the geometrical meaning of all the concepts, even more so as it was not clear at this point that the application of the ADC would be fruitful. Einstein and Grossmann probably first wanted to be sure that the approach could solve a substantive problem, say, yielding the right classical limit, to confirm that they were on the right track. When this did not work, the approach was, to a certain extent, abandoned, and it would have been a waste of time to work out what the geometrical meaning of, say, the Riemann tensor is.

The disadvantages of working with an uninterpreted calculus generated some persistent problems in the genesis of GR. It simply was not clear what the mathematics was describing, or what aspects of it had to be interpreted realistically. The interpretation of the significance of manifolds, and their relation to the metric, was one major stumbling block.

### 6.10.4 Open Problems and Possible Extensions

Finally, here are some open problems and possibilities for extending the above study.

- A systematic evaluation of sources, other than the Entwurf and the papers cited therein, especially the Zurich notebook and further texts from pure mathematics, could deepen our understanding of the application of tensor calculus in GR.

- Our understanding of some details of algebraic invariant theory are still incomplete. In particular, we would like to understand better whether, at this early stage, it would have been possible for Grossmann to find the Einstein tensor. This is related to the question of how to interpret the "complete set of differential invariants" locution by Grossmann, Ricci & Levi-Cività, and Christoffel.

- The historical question as to what Grossmann contributed to the formulation of the question that led to the application of the ADC is still open. It is probably impossible to answer this question without further historical sources.

# Chapter 7

# Introducing the Inferential Conception

## 7.1 Introduction

In this chapter, we introduce and discuss a recent account of the applicability of mathematics to the world, the Inferential Conception (IC) proposed in Bueno and Colyvan (2011).[1] The chapter has three objectives. First, we give a short exposition of the IC, in section 7.2. Then, in section 7.3, we offer some critical remarks on the account, and discuss potential philosophical objections. Third, we propose some extensions of the IC in section 7.4, preparing the ground for the application of the IC to our case study in chapter 8. We conclude in section 7.5.

## 7.2 Mapping Account and Inferential Conception

Our account of the IC in this section follows the presentation in Bueno and Colyvan (2011). First, we introduce the "default view" of the application of mathematics, the so-called *mapping account.* Then we discuss four problems of this account. Third, we introduce the IC and show how this account solves the problems of the mapping account.

### 7.2.1 The Mapping Account

Bueno & Colyvan address the question of the role of mathematics in application. They use a familiar picture of applying mathematics as a foil for their own account. On this "default view", mathematics helps us in application, by representing empirical structures in mathematical form; we can then learn about the world, by examining this mathematical representation. The application relation is established via a structure-preserving mapping,

[1] This chapter is based on joint work with Tilman Sauer.

which connects the mathematical structure with relevant parts of the world. The mapping account has been spelled out in detail in Pincock (2004).

A city street map is a useful illustration of the mapping account. A city street map represents parts of the structure of a city by mirroring the street system and buildings of the city in some detail. A map will usually leave out some information such as slope, or even distances. However, there should be some correspondence (mapping) between parts of the street map and parts of the city – most importantly, it should represent how various parts of the city are connected. This information can then be read off the map, and therein lies its usefulness.

### 7.2.2 Problems of the Mapping Account

Bueno & Colyvan find the mapping account to be largely correct. However, they argue that it cannot be a complete story of how mathematics is applied, and why it is useful in application. They identify four problems with the mapping account, and claim that the IC fares better on all four counts.

The first problem is the *assumed structure problem*. Normally, when we speak of a structure-preserving mapping between two domains, we assume that both domains are equipped with a set of objects and relations (properties) between these objects, or, more generally, that there is some sort of structure present in both domains. Therefore, if we want to account for the applicability of mathematics based on the mapping account, we have to assume that some sort of structure is in the world that can be preserved by the mapping. However, there is simply no guarantee that the world is conveniently structured in this way. We have to take it for granted that there is some meaningful way of discerning and using the structure of the world.[2]

The second problem, which we dub the *choice of mapping problem*, arises if we try to be more precise about what kinds of mapping are acceptable, and say more about how to find acceptable mappings in the first place. On the mapping account, we expect the mapping to preserve structure, so it should be a homomorphism of some sort. Additionally, the mapping can be surjective, injective, or both.[3]

However, the idea that we can capture all that is essential to application with just one structure-preserving mapping is problematic. At times, we do not want to take into account either all the available structure in the world (think of the city street map example above), or all the "surplus" mathematical structure (think of physically meaningless solutions to some

[2]Bueno & Colyvan write that this might not be a problem in the case of the city street map. However, we think that even in this simple case, we have to explain carefully how to interpret the assumed structure; see the discussion of the related Königsberg case in chapter 2.

[3]A mapping $f : X \to Y$ is surjective if $f(X) = Y$, injective if, for all $a, b \in X$, if $f(a) \neq f(b)$, then $a \neq b$. A mapping that is both injective and surjective is bijective, i.e. there is a one-one correspondences between $X$ and $Y$.

mathematical equation). What are the criteria for selecting the relevant parts, both in the mathematical domain and in the world? Bueno & Colyvan think that any account of applying mathematics that is silent on this issue should be considered incomplete.

The third problem is *idealization*. If a mathematical structure is idealized, we know that it does not faithfully represent an actual structure in the world. An example is when we use a continuous mathematical structure to represent quantities that are really discrete (think of the continuous Lotka-Volterra equations that represent discrete predator and prey populations; see chapter 3). In these cases, a mapping between the two domains is impossible; there can only be a mapping between a mathematical and a possible, but non-actual, empirical structure. An account of how mathematics is applied should say something on this issue.

The fourth problem has to do with *explanatory contributions* of mathematics. If all that mathematics does is to represent certain aspects of an empirical system, then there is no place left for an explanatory role of mathematics, because every explanation based on the representation could also be given on the basis of the features of the system represented by the mathematical structure. This, however, is in tension with examples where mathematics has a *prima facie* explanatory role, be it unificatory or otherwise.

### 7.2.3 Introducing the Inferential Conception

Next, Bueno & Colyvan introduce the Inferential Conception (IC). One of the motivations behind the IC is to solve the above problems of the mapping account. The IC breaks down the process of applying mathematics into three steps:

1. In the *immersion step*, we specify a mapping from the relevant aspects of the empirical domain to a mathematical structure.

2. In the *derivation step*, also called deduction step, we draw inferences from the immersed mathematical structure. This step is purely mathematical.

3. In the *interpretation step*, the consequences found in the derivation step are mapped back to the empirical domain, that is, we interpret the results of our mathematical investigation. The mapping we use in the interpretation step is not necessarily the inverse of the immersion mapping; it need not even be of the same kind.

It is one of the main innovations of the IC to distinguish the immersion and the interpretation step; this feature is absent in the mapping account. Mappings play a more flexible role in the IC than in the mapping account.

Distinguishing the immersion and interpretation steps gives the IC a distinctively dynamical flavor – the suggestion is that by going back and forth between mathematics and the world, we can gradually refine the mathematical description, and also discover new empirical phenomena. This dynamic is a second important feature of the inferential conception.

The emphasis on inferences is an important feature of the IC. Bueno & Colyvan describe the usefulness of mathematics in application emphasized by the IC as follows:

> [B]y embedding certain features of the empirical world into a mathematical structure, it is possible to obtain inferences that would otherwise be extraordinarily hard (if not impossible) to obtain. (Ibid., p. 352)

Facilitating inferences is not the only role of mathematics in application. Bueno & Colyvan maintain that most, if not all, other roles of mathematics in application, especially unification, novel prediction, and explanation, are ultimately tied to its inferential role.

Bueno & Colyvan do not discuss the distinctive roles that the *mathematical domain* plays in application. They only note that often, the mathematical formalism has an intended empirical interpretation.

The discussion of the *empirical domain*, Bueno & Colyvan point out that the empirical domain is, to a certain degree, dependent on mathematics: often it is impossible to give a mathematics-free description of the empirical domain. They nevertheless adopt a metaphysical reading of the empirical domain – it is real, empirical structure, and does not depend on the possibility of describing it in non-mathematical terms. A second issue is the metaphysical nature of the empirical domain. Bueno & Colyvan emphasize that the structure in the empirical domain can, but need not be, causal structure. They mention examples proposed by Batterman (2002), suggesting that abstraction from causal details can be important for the application of mathematics, e.g. for explanatory purposes. Finally, the inferential conception encompasses the possibility of applying mathematics to itself; in these cases, the domain of application is not empirical. Bueno & Colyvan think that a clean separation between empirical and mathematical domains is not as important as the existence of immersion and interpretation mappings.

### 7.2.4 Advantages of the Inferential Conception

Once they have introduced this scheme, Bueno & Colyvan explain how the IC is supposed to solve the four problems of the mapping account. Our explanations here are rather short, because we focus on those parts of the IC we will use later.

First, to the *choice of mapping problem*. This problem has two aspects. First, there is the problem that there might be surplus structure in both domains. The IC is more flexible in that it is not restricted to using just one mapping and its inverse, and can thereby avoid the formal problem from surplus structure. The emphasis is not on a structural correspondence, but on facilitating inferences, which is a contextual matter to a large degree. Second, the mapping account is silent on how to choose the right mathematical setting (and, by extension, the right mapping), and assess the adequacy of that choice. The IC can avoid this problem by going back and forth between mathematical and empirical domains, which makes it easier to gradually revise the relation between the empirical domain and a mathematical structure.

Second, the inferential conception can help to solve the *assumed structure problem*: We can start from a tentative assumed structure in the world, and gradually revise the structure, after our inferential investigations, and by choosing a new interpretation mapping; Bueno & Colyvan write that there is no need to "formally revise" the initial assumed structure, as the interpretation mapping need not be the the inverse of the immersion mapping. Once more, the back-and-forth between empirical and mathematical domains and the emphasis on inferences are key.

Third, Bueno & Colyvan introduce the so-called *partial structures* framework, to deal with the problem of *idealization*. The idea behind this framework is to allow mappings and structures to be only partially defined. Bueno & Colyvan present this framework in some detail, and explain its workings using an example from economics. We will not explore this part of Bueno & Colyvan's solution. This is harmless because the partial structures framework is largely orthogonal to the overall scheme of the IC, and can be assessed independently.

Fourth, Bueno & Colyvan tie the theoretical virtues of *unification, novel predictions and explanations* to the inferential conception, by emphasizing the role of inferences in unifying, predicting and explaining. They think that these aspects depend on contextual factors, and they mention that mathematics might help to give us epistemic access to theories, by highlighting inferential patterns. The problem of the tension between the representational and explanatory roles of mathematics is not explicitly addressed, but in their discussion of examples by Batterman (2002), they leave open the possibility for an explanatory role of mathematics that goes beyond pure representation.

## 7.3 Discussion of the Inferential Conception

In this section, we take a closer look at those aspects of the IC that are relevant for our purposes. We will point out some weak points of the IC, and explain how the account might be amended.

### 7.3.1 Empirical Domain

On the IC, the empirical part of the application process has, at least implicitly, two aspects. The first is the empirical domain, i.e. the domain of application insofar as mathematics is not applied to itself.

Bueno & Colyvan do not give a definite answer to the metaphysics of the empirical domain. They think that characterizing the empirical domain in terms of causality, is possible, but not necessary, for two reasons: Firstly, there are cases where the application of mathematics works because we leave out causal details; Bueno & Colyvan mention the Euler strut discussed in Batterman (2002) as an example. Secondly, they want to leave open the possibility that the domain of application is itself purely mathematical. Bueno & Colyvan even suggest that empirical and mathematical domains are not clearly separable: "We are not assuming, in the immersion and the interpretation steps, that the empirical set up and the mathematical structures are completely distinguished components" (Bueno and Colyvan, 2011, p. 354).

We agree with Bueno & Colyvan on the first point – to a certain extent. An account of applying mathematics should leave open the possibility of non-causal explanations. A second concern about a purely causal characterization of the empirical domain is that it could be problematic to interpret geometrical structure as causal. Geometry and causality are certainly intimately connected, but it is not clear that geometry, in itself, is causally relevant (think of higher level explanations using geometry, not space-time theories such as GR).

We are more sceptical about the second reason for not characterizing the empirical domain as causal. We think that the application of mathematics within mathematics, and empirical application, should be carefully distinguished. We will discuss the reasons for distinguishing these two cases in section 7.3.4 below.

### 7.3.2 Assumed Structure

The second aspect of the empirical part of the application process is the assumed structure. This is "something like a pre-theoretic structure of the world (or at least a pre-modeling structure of the world)." (Ibid., p. 347) As the name "assumed structure" suggests, this is not structure that we should take metaphysically seriously. Rather, it is "tentative" structure that can be revised once we have brought mathematics into play: "the assumed structure is the structure the modeling exercise assumes to be present in the [...] empirical setup [...] the interpretation step of the process will deliver the final structure of the empirical setup [...]" (Ibid., p. 357). Bueno & Colyvan point out that the assumed structure need not be mathematics-free: "It will be hard to even talk about the empirical setup in question without leaning heavily on the mathematical structure, prior to the immersion step."

(Ibid., p. 354). According to this picture, the assumed structure can already be mathematized. The immersion maps the mathematically characterized assumed structure to the mathematical domain.

Several philosophical objections can be raised against the IC because it relies on the concept of assumed structure.

A first objection[4] is that the assumed structure problem is an insurmountable difficulty for the IC along with all accounts that rely on mappings. We should conceive of the immersion and interpretation mappings as mathematical objects. Therefore, both domain and codomian have to be mathematical as well. But then a mapping cannot account for the application of mathematics to the world. All we have is a mapping from an assumed mathematical structure, which is not really in the world, to mathematics. The IC cannot possibly explain how mathematics can be applied to the world; it can only explain how mathematics can be applied to some other mathematical domain. The IC is circular.

We think that this problem can be dissolved. We agree that, in some cases, including the case study we are about to consider, the assumed structure is given in mathematical form. However, this does not mean that the structure itself is mathematical. Rather, the mathematics represents, possibly in an indirect manner, empirical facts. The immersion mapping establishes a correspondence between these empirical facts and a mathematical structure, not between the mathematical representation of the empirical facts and another mathematical structure. We will spell this out in more detail below.

Also, it has to be appreciated that this objection does not only apply to the IC, but to a whole class of accounts of how mathematics is applied. The objection can be raised once we accept that, first, there is a clear separation between mathematical and physical domain, and b) that some kind of mapping, however indirect, between the two domains has to exist. Both of these assumptions seem very reasonable. Thus, although the objection appears *prima facie* attractive, the price you pay for accepting it is quite high.

The objection nevertheless raises an important issue: we have to explain the value of basing an account of applying mathematics on mappings, because it seems that mappings can only connect mathematical domains. We will make an effort to be clear on the question as to what the starting point of the application process is, viz. how the assumed structure is constituted, and also to explain why the assumed structure is not (purely) mathematical.

A second objection[5] is to deny that there is a problem here at all, at least from the perspective of scientific realism: *of course* can we find objects, relations, structures in the world that are independent of mathematics. The

[4]We thank members of the audience at the Workshop "Metaphysics of science: objects, relations and structures" of October 2012 in Lausanne for this objection.

[5]We thank Matthias Egg for raising a form of this objection in the philosophy of science research seminar in the fall of 2012 in Lausanne.

assumed structure problem is a red herring.

We think that this objection only has traction if we adopt a metaphysical reading of the assumed structure, i.e. if we presuppose that it is unproblematic to interpret the assumed structure as real, empirical structure that is mapped to mathematics. However, if it is the goal of, say, a new theory of gravitation to unveil the real structure of the world, but this theory is not yet available, how can we take that very structure as a starting point of our investigation? Saying that the assumed structure is just the structure in the world seems like begging the question.

When Bueno & Colyvan write about the assumed structure as the "relevant bits of the empirical world", this should not be read metaphysically. We should interpret the assumed structure as the reasonable starting point of the *process* of applying mathematics. It is not to be confused with the *result* of applying mathematics.

*Prima facie*, there are two separate issues here: the metaphysical question as to what is the nature of the structure we are trying to capture mathematically, and the epistemological question as to how we come to know this structure according to the IC, viz. by going back and forth between mathematics and the world, and thereby refining our mathematical description.

However, we think that these issues cannot be separated in a clear-cut way. A theory which is the final product of a mathematical description of the structure of the world is a product of that discovery, and thereby closely tied to the reliability of the process by which we discover the theory.

A third problem with the assumed structure is the choice of an appropriate starting point. According to the IC, we start with a tentative assumed structure, which can be gradually revised by going back and forth between the empirical domain and mathematics. However, the IC does not have anything to say about how the initial assumed structure should be selected. The concern could be that if we choose a bad starting point, all the work of going back and forth could lead to a bad outcome.

We see three possibilities for dealing with this problem. First, we could formulate a systematic account of how to select an appropriate assumed structure. Second, we could come up with an argument showing that the "revision process" of the IC can "wash out" bad starting points. The third possibility is to deny this is a problem at all: The IC is an account of how mathematics is applied, not an account of how mathematics is applied successfully. The IC leaves open the possibility that something goes wrong in the application process, and this could well be due to a bad starting point.

Currently, we do not know how to substantiate ideas that would support options one and two. For the time being, we endorse option three. If it were possible to say more on the first two options, the IC would be an account of how to apply mathematics successfully.

Summing up, the following picture of the assumed structure in the IC emerges: Assumed structure is tentative empirical structure, possibly represented in mathematical form. It is not the empirically adequate structure of the world, and we should therefore not take it metaphysically seriously. The goal of the process described by the IC is to revise and refine the initial assumed structure. One of our goals in the case study will be to identify the assumed structure in the beginning, and to track the revision process.

### 7.3.3 Mathematical Domain

Bueno & Colyvan argue that the inferential conception does not have to settle for a position in the metaphysics of mathematics. We will follow them in this respect and not enter into these debates.

The focus of the IC (and of the mapping account) is on finding mappings from one domain to another. This picture suggests that the domains are, to a certain extent, static objects. By adopting this perspective, we run the risk of overlooking that the domains themselves, and the mathematical domain in particular, are dynamical. We are not claiming that such a dynamic picture of domains is incompatible with the IC; rather, it is an extension of the view.

We will see in our case study that the mathematical domain is indeed a dynamical entity. Note that these aspects of the mathematical domain counteract, to some extent, the thesis of the "unreasonable effectiveness" of mathematics: the more we can attribute the successful application of mathematics, either to an effort prior to application, or to adaptation during application, the less the effectiveness is miraculous, or unreasonable.

Finally, it will prove fruitful to be specific about the mathematical theory that is applied. We should prevent our picture from becoming too coarse-grained by lumping all mathematical theories under one header. What is more, even if a particular kind of immersion or interpretation mapping is specified, this need not determine the mathematical theory that is applied, especially if it is possible to interpret one mathematical theory in terms of another (think of a calculus with and without "intended interpretation"). This is compatible with Bueno & Colyvan's idea of multi-stage application processes, where one of the stages is purely mathematical.

### 7.3.4 Separating Empirical and Mathematical Domain

Bueno & Colyvan claim that the IC can encompass both application to the world, as well as the application of one mathematical theory to another. While we agree that both these notions are important, and worth exploring, we maintain that there are good reasons for working them out separately, for two reasons.

First, a big part of the puzzle surrounding the application of mathematics stems from the fact that it is unclear how mathematics can help us solve empirical problems. On the other hand, the question as to how the application of one mathematical theory to another works, while also worth asking, is not nearly as puzzling. This suggests that these are at least two clearly separate problems; the applicability of mathematics to the world is more pressing.

Second, it has been argued in the literature that attempts at drawing a clear distinction between mathematics and the empirical domain, say, in terms of the abstract-concrete distinction, are in vain, see Ladyman and Ross (2007, pp. 159), or even that the distinction between mathematics and the world is blurred, see e.g. French (2000). While it may be true that there is no wholesale solution to the problem of distinguishing between pure and applied mathematics, this does not imply that the distinction can therefore be dismissed easily, or that no solutions exist in each case. Quite to the contrary, separating the representation from what is represented is at the core of many philosophical debates, especially in the philosophy of physics.

One example is the question as to how to interpret the wave-function in quantum mechanics – are some aspects of these objects mathematical artefacts, and if yes, which ones? In GR, there is, first, the problem of the correct interpretation of coordinates, which plagued the genesis of GR, and, secondly, the problems arising from the Hole Argument, which turns on the question as to what the correct mathematical representation of space-time points is. All these examples suggest that giving up on a clear separation of the two domains would amount to dismissing all these problems. This is not a viable option.

### 7.3.5 Mapping Selection Problem and Dynamics

Bueno & Colyvan propose solving the mapping selection problem by, firstly, distinguishing immersion and interpretation mapping, which makes the correspondence between domains more flexible. On the mapping account, working with just one homomorphism is responsible for the problem of surplus structure. The second part of the solution is the suggestion that a back-and-forth between empirical and mathematical domains will help us find the right kind of mapping, or structure, that is preserved.

We agree with Bueno & Colyvan that the IC solves the (formal) problem with surplus structure, and therefore will not further discuss this issue. We will focus on the second aspect of the solution, which we call the dynamical aspect of the IC. By dynamics, we mean the interaction of the IC's components such as mappings and domains, and also regularities between these components. We think that Bueno & Colyvan's idea that there exists a back-and-forth between empirical and mathematical structure, which helps us choose both the right assumed structure and the right mathematical setting, is an interesting idea that should be spelled out in more detail. The picture of gradual refinement is an attractive picture of application, but there is no guarantee that this process will be successful, or lead to a good equilibrium point. The IC is a theory of the application of mathematics, not a theory of successful application. What the IC can do is to provide us with diagnostic tools if an application goes wrong, i.e. if there is a mismatch between some empirical target structure and a mathematical structure. We will say more on this issue in the next section.

A second set of questions concerns the starting point of the dynamics. What prompts the search for a new mathematical domain? This question is of special interest in cases where there is already a mathematical structure in place. What kind of consideration leads scientists to question, or abandon, an old framework, and start to search for a new mathematical domain? What are the guiding principles if a scientist does not yet know what kind of mathematical framework he needs in the first place? We will turn to these questions in our case study.

Then, not all components of the IC are independent. The mathematical domain determines, to a certain degree, the immersion and interpretation mapping, as the mathematical structure has to be able to mirror the structure preserved by the mapping. What is more, there is not just one kind of structure that can be preserved, we have to specify the nature of the structure – graph structure, ordinal structure, metric structure, and so on. This leaves open the possibility of surplus structure on either side of the mapping.

### 7.3.6 Idealization (Partial Structures)

Bueno & Colyvan think that the mapping account is incomplete because it cannot accommodate idealizations. They propose solving this problem by augmenting the IC with the partial structures framework.

We agree that idealizations are a pressing problem for any account of the applicability of mathematics. However, we are sceptical that the partial structures account can address the problem. Here we offer some reasons for our critical stance towards partial structures. However, we will not enter into a sustained critical discussion. Our criticism does not invalidate the IC, as the two frameworks are largely independent.[6]

First, partial mappings are not necessary for solving the problem of surplus structure in the domain. It is sufficient to suitably restrict the domain of a normal mapping such that all and only those objects that are not superfluous are mapped.

Second, we think that the partial structures account might be too permissive. For example, it is possible to have a mapping between empirical structure and mathematics, even if there is a total mismatch between the two structures. The domain of the partial mapping can simply be assumed to be empty. The partial structures account runs the risk of being trivial, because it allows for structural correspondence between any two structures.

Third, Bueno & Colyvan introduce a case study from economics to illustrate the force of the partial structures account. The partial structures

[6]See e.g. Pincock (2005) for further criticism of the partial structures framework.

account is attractive because it makes it possible to take vague ideas about a partial match between empirical and mathematical structures and spell them out in a formal framework. However, the discussion in the paper remains at an informal level. The authors do not really exploit the strength of this account. It would be desirable to see a case study where the formal apparatus of partial structures is really put to work.

Finally, we think that the defenders of the partial structures account should demonstrate that their account is able to solve some of the classical problems of idealization, for example by spelling out the important distinction between "harmless" idealizations, i.e. false modeling assumptions that do not prevent a model from being useful, or even admit of realistic interpretation, and "essentially idealized" models, which pose a threat to realism.[7]

## 7.4 Extending the IC

In this section, we introduce some extensions of the IC. These extensions are designed to accommodate a distinctively historical approach to applying mathematics, i.e. to mirror not only the result of application, but the historical process leading to mathematically formulated theories of empirical phenomena. The role of the IC, and its extensions, is that of a conceptual framework providing us with tools to better understand historical cases. It is not meant to be a wholesale account covering every aspect of the application of mathematics.

Bueno & Colyvan at least implicitly construe immersion, deduction, and interpretation steps as temporal; otherwise, talk of a gradual revision process, based on going back an forth between mathematics and the empirical domain, would not make sense. The temporal succession between these steps need not be temporally ordered, as beginning with immersion, followed by deduction and then interpretation, in all cases. For example, at times, we will check possible deductions before working out the immersion step.

The smallest historical unit that is iterated in temporal succession consists of immersion, deduction and interpretation steps. We call such a unit a *cycle*. A cycle is one round of going back-and-forth between mathematics and the world, one round of assimilating a mathematical theory and a particular empirical structure. The starting point of a cycle is the initial assumed structure, the end point is the revised assumed structure.

We conceive of the IC as an account of application of mathematics, not an account of successful application of mathematics. Accordingly, there are two possible outcomes for completed cycles. The application can be successful, i.e. the scientist finds that mathematics and empirical structure match as expected. We call this a *closed cycle*. On the other hand, the application can

[7]See e.g. Uskali Mäki's work on idealized models in economics. Lehtinen et al. (2012) is a useful recent survey.

fail in the eyes of the scientist; we call this an *open cycle*. The two possible cycle outcomes, closed and open cycles, trigger different dynamics. We first discuss the dynamics of open cycles.

If the outcome is an open cycle, i.e. the scientist finds that there is a mismatch between the mathematically derived result and the empirical domain, a process of reflection is set in motion: the failed application can be due to any one of the components of the IC, and we can use the components as a diagnostic tool to analyze the open cycle, and potentially fix the problem. Of course, the problem can be with more than one of the components.

First, we can attribute the failure to a bad initial assumed structure. Maybe the empirical phenomena we took as a starting point of our modeling exercise have to be investigated more closely, and more data have to be collected. Maybe it is even necessary to reconceptualize our description of the empirical phenomena.

Second, the failure can be due to the immersion step, or the mathematical theory we use. The framework can lack the expressive power for a task, it can be insufficiently understood, or it can exhibit internal difficulties, or even inconsistencies. In this case, the problem can be fixed, either by revising, or further exploring, the mathematical framework, or it could be given up entirely.

Third, there could be a failure in the deduction step. The scientist could commit a mistake; certain inferences can be hard, or even impossible, to reach, so that it is hard to reconnect the findings of the deduction step with the world in the interpretation. In this case, the scientist could make an additional assumption to facilitate the deduction, search for a different framework, or she could explore alternative routes of deduction.

Finally, there could be a failure in the interpretation step. For example, it could be unclear whether a solution to some equations has to be taken seriously, or if it is just a mathematical artifact.

In the case of a closed cycle, a successful application step, the process of application is not over. Usually, a closed cycle only means that the scientist has successfully derived some consequences within mathematics that have a meaningful empirical interpretation, not that there is a full match between the mathematical and empirical structure. The further goal will normally be to consolidate the closed cycle. Again, improvements are possible in all components: the scientist can send the revised assumed structure back to mathematics in the second cycle, or even widen the scope of phenomena in the assumed structure, and check whether the cycle can still be closed; she can work out further deductions and interpretations from the original immersed structure; the validity of the deductions can still be checked; and, the interpretation can suggest further empirical investigations, if it yields novel predictions that have not been anticipated, and so on.

In sum, a historical reading of the "equilibrium process", the mutual adaptation of mathematics and the world, in terms of an iterated game of cycles, which can either be successful or fail, seems to us to provide a useful analytic tool that will help us better understand the genesis of mathematically formulated empirical theories, and the systematic interplay between mathematics and the world.

## 7.5 Summary

In this section, we sum up the essential points about the extended IC that we will put to work in the historical case study in the next chapter. We can distinguish the components of the IC and the dynamics of these components; both come with a set of systematic questions.

**Empirical Domain and Assumed Structure:** The empirical domain consists of those phenomena that are the subject of theorizing. The assumed structure, on the other hand, is the structure we assume to be in, or part of, the empirical domain, such that the process of application can get started. The assumed structure need not be the real structure of the world; it is subject to revision in the process of application. It is a “modeling choice”. *Questions:* What are the empirical domain and the assumed structure? In what form is the assumed structure given (specificity)? Is the picture of initial and refined assumed structure adequate?

**Mathematical Domain and Deduction Step:** This is the mathematical structure, or theory, that is applied. Deduction is the step of drawing inferences from the mathematical structure within mathematics. *Questions:* Can we confirm the picture of the mathematical domain as a dynamical entity, in that a) there is a tendency within mathematics to prepare or anticipate application, and b) the mathematics is adapted, changed, expanded during the process of application? What is the role of interpretation of one mathematical theory in terms of another, or “layered application”? What is the role of notation in the deduction step? How does the possibility of drawing inferences influence the choice of mathematical theory?

**Immersion and Interpretation Step:** Immersion is the mapping from the initial assumed structure into the mathematical domain; whereas interpretation is the mapping of the inferential results back into the empirical domain, not necessarily the inverse of the immersion mapping, or even of the same type. *Questions:* Do we have instances where the two come apart in kind? What is the systematic upshot of the distinction in the historical case?

**Dynamics, Back-and-Forth, Open and Closed Cycles:** The idea behind the dynamics of the IC is that by going back and forth between empirical and mathematical domains, we can revise the initial assumed structure after repeated cycles consisting of immersion, deduction, and interpretation. This can, but need not, lead to an equilibrium between empirical structure and mathematical representation. If we have a closed or an open cycle, depending on whether the application is successful or not, different kinds of "reflection" on the application process are triggered.

*Tasks:* Find instances for these theoretical ideas, especially for closed and open cycles, and the reflection process they trigger. Refine the picture suggested here accordingly.

# Chapter 8

# Applying the IC to GR: Three Episodes

## 8.1 Introduction

In this chapter, we put the Inferential Conception (IC) to work in our historical case study, the genesis of GR.[1] We will analyze three historical episodes using the conceptual apparatus provided by the IC. This, in turn, will help us to refine the account.

In section 8.2, we investigate how the starting point of the application process, the "assumed structure", is chosen. We will clarify the status of the starting point of the application process, and discuss the trigger of the application process. Then we analyze two small application cycles that led to revisions of the initial assumed structure.

In section 8.3, we examine how the application of "new" mathematics – the application of the Absolute Differential Calculus (ADC) to gravitational theory – meshes with the IC. We describe how the mathematical part of the Entwurf is shaped by the application process. Our focus is on the application cycle that led to the "discovery", and application, of the ADC, the quest for generally covariant differential operators.

In section 8.4, we will take a closer look at two of Einstein's failed attempts to find a suitable differential operator for the field equations, and apply the conceptual tools provided by the IC to better understand why he erroneously rejected the Ricci tensor and the November tensor in the Zurich Notebook. We conclude in section 8.5.

[1]This chapter is based on joint work with Tilman Sauer.

## 8.2 Episode One: Choosing a Starting Point

In this section, we identify and discuss the initial assumed structure, the first component of an application cycle, at the beginning of the genesis of GR. We are not only interested in the historical starting point of the "modeling exercise" of GR, but also in the systematic lessons to be learned from the episode. After proposing a candidate for the initial assumed structure in GR, we track the first rounds of application cycles, leading to the first revisions of the initial assumed structure.

### 8.2.1 What Triggered the Application Dynamics?

One of the problems prompting the search for a new theory of gravitation was the conflict between special relativity (SR), which postulates a finite, constant speed of light $c$, and Newtonian Gravitational Theory (NGT), which postulates an instantaneous propagation of gravitational effects; see section 5.2.1. These two physical domains were inconsistent and had to be reconciled.

While the inconsistency is rooted in physics, it carries over to a formal inconsistency in their respective mathematical formulations: NGT does not conform to the formal requirement imposed by SR, Lorentz covariance. Therefore, the mathematical formulation of the theories had to be adapted as well. This set the process of the application of mathematics in motion. However, it was not yet clear whether radical changes in the mathematics, or the application of a new kind of mathematics were necessary, or whether a more conservative revision of NGT was sufficient.

### 8.2.2 The Initial Assumed Structure in GR

What was the initial assumed structure at the very beginning of the genesis of GR? Recall, from section 7.3.2, that the initial assumed structure is some aspect of the empirical domain, which is mapped into mathematics at the beginning of the application process. The assumed structure need not be real empirical structure, that is, we need not take it metaphysically seriously. Also, the assumed structure can be represented in mathematical form. It has to be kept in mind that the assumed structure is what is represented by the mathematics, not the representation itself. We will have to spell out what this means in each case.

Historically speaking, the starting point of Einstein's investigations were the two theories that triggered the search for a relativistic theory of gravitation, SR on the one hand, and NGT on the other. We think that these two theories, or the empirical structure they describe, are the assumed structure for the modeling exercise of GR. Both are formulated mathematically. Therefore, it is vital to specify what empirical phenomena they describe.

SR can be interpreted as a restriction on physical theories, such as electrodynamics, requiring that their laws are Lorentz covariant. NGT, on the other hand, describes gravitational phenomena – the motion of massive bodies in space due to the force of gravitation, such as the motion of planets around the sun, and falling bodies on earth.

The case of NGT is a little more tricky. First, while NGT is well confirmed for small velocities, this is not true for velocities approaching the speed of light, where the inconsistency with SR becomes apparent. Second, some isolated gravitational phenomena, in particular the anomalous perihelion advance of Mercury, are incompatible with NGT. It was known for quite some time that NGT could not account for this anomaly, which is also a gravitational phenomenon.

These two lines of conflict were not treated equally. While the first shortcoming was at the center of Einstein's attention – the conflict between SR and NGT triggered the search for GR, and it was one of the main goals of the new theory to remove the inconsistency between the two theories – the second class of phenomena did not enter into the construction process of GR. It is not part of the assumed structure.

In sum, the physical domain of gravitational phenomena that Einstein was dealing with was given in the form of NGT. He took it to be an empirically well validated theory that was to be reproduced by the successor theory, keeping in mind, however, that two kinds of inconsistencies, that with SR, and that with isolated phenomena, had to be removed. Thus, it was obvious from the very beginning of the genesis of GR that the initial assumed structure would not stand as it was, but that, ultimately, it would have to be revised. Also, not all relevant empirical phenomena were part of the assumed structure.

### 8.2.3 Assumed Structure: Systematic Significance

This account of the initial assumed structure in the genesis of GR is historically accurate. But what is its systematic significance? Is Einstein's choice of initial assumed structure a mere historical contingency?

We think that there are systematic reasons for Einstein's choice of initial assumed structure; they can be traced back to the heuristic correspondence principle, one of the guiding requirements shaping the search for GR (see section 5.2.4). Recall that, according to the correspondence principle, the new theory should reproduce NGT in a special-relativistic, weak-field limit, such that the empirical knowledge embodied in NGT was sure to be recovered in this limit. Everything that could be adequately represented by NGT could be adequately represented by a generalized theory as well, provided that the generalized theory reduced to the Newtonian limit properly.

The heuristic requirement of a Newtonian limit therefore fulfilled a function of utmost importance: it guaranteed that the new theory would be empirically adequate, at least to the same extent as NGT, and the empirical adequacy of the new theory would not have to be demonstrated from scratch.

However, the role of the correspondence principle is not limited to empirical adequacy, it also plays a theoretical role. By specifying how we approach the Newtonian limit, we also provide information about the necessary background assumptions that are required for the validity of the Newtonian theory, i.e., we obtain information about the possible reasons for empirical failures of NGT, and we get information about the order of magnitude of numerical errors that the old theory generated in the prediction of, say, the anomalous perihelion advance of Mercury.

As it turned out later, the Newtonian limit had a third, equally important, role. By establishing the connection with NGT, it was possible to interpret mathematical objects introduced by the new mathematical theory. For example, the $g_{00}$-component of the metric is connected with the Newtonian scalar potential. The Newtonian limit is also the standard way of determining the physical dimension, and numerical value, of the arbitrary constant entering the gravitational field equations.

In sum, the systematic significance for this choice of initial assumed structure can be found in the heuristic correspondence principle. The correspondence principle played at least three roles. It made sure that most known gravitational phenomena are saved, if the new theory reduces to the predecessor theories in a suitable limit; it made it possible to explain, after the fact, why NGT failed in some instances, and it helped in determining constants in the new theory.

### 8.2.4 But is it Application?

It could be questioned whether, at least at this stage of the story, we are dealing with a genuine case of the application of mathematics. The use of the correspondence principle could suggest that this a case of inter-theoretic reduction, not of the application of mathematics. We do not take empirical phenomena as a starting point, but the initial assumed structure already comes in mathematical form.

To this we reply that, first, the correspondence principle plays multiple roles here; one of them is to save the phenomena of NGT. Building the old theory into the new one is not only a theoretical exercise, it serves the purpose of connecting the new theory with the empirical structure described by the old theory. In this respect, the mathematics of the new theory is applied to a (known) empirical structure.

Secondly, once physical theories have reached a certain degree of sophistication, it is to be expected that the empirical phenomena will be represented in mathematical form. The genesis of GR is an example of this. We do not start from scratch, but, as suggested by Bueno and Colyvan (2011), we pick up the construction of the theory mid-stream. We have already gone through several cycles of immersion, deduction, and interpretation, and what we see here are further application cycles.

### 8.2.5 Refining SR and NGT: First Steps

We have now characterized the initial assumed structure of GR at a rather general level. Now we will put the IC to work. We will sketch, for some instances, how the initial assumed structure was refined before the application of the ADC. The picture will also get more detailed as we describe the application of particular mathematical theories. We believe that it is necessary to describe the process of application at a rather fine-grained level, as crucial breakthroughs in the episode hinge on the use of particular formulations of the physical theories.

#### The Poisson Equation

First, what form of NGT did Einstein adopt? The starting point of the modeling exercise in classical mechanics, the assumed structure, was the Poisson equation, the classical, non-relativistic field equation of gravitation; see 5.2.2.[2] One of the reasons why the Poisson formulation of gravitation was preferred over Newton's force law was the former's greater mathematical simplicity; it is a scalar equation. The main reason why it was taken as a starting point in the genesis of GR was that it fitted better with the Lorentz model approach, on which gravitation, and other forces, are mediated by a field, which, in turn is determined by the sources. Thus, Einstein's main reason for adopting the Poisson equation was not mathematical, but physical: he wanted to find a relativistic field equation; the goal of the enterprise was to avoid Newton's action-at-a-distance force law.

#### Einstein vs. Abraham: Adopting Minkowski

Second, let us have a closer look at the form of SR. Prior to 1912, Einstein was content with the "pedestrian" account of SR he had developed in his 1905 paper on SR. This account did not make use of group-theoretic notions or the line element provided by Minkowski's formulation of SR, but rather used explicit coordinate transformations. As late as 1911, he quipped that he did not understand "modern" formulations of SR, based on the Minkowski formalism, which was discussed in a textbook on SR written by Max von Laue (see Norton (2000, p. 141)).

It was only after Einstein was confronted with a rival, special-relativistic approach to gravitation, by Max Abraham, that he changed his mind and

[2] Historically, the first attempts to formulate classical mechanics in a Lorentz covariant manner took Newton's law of universal gravitation as a starting point. It is not clear whether Einstein first took this route as well.

adopted the more powerful approach by Minkowski.[3] Abraham's theory of gravitation was a straightforward generalization of the Poisson equation, based on the Minkowski line element. Einstein must have been surprised when Abraham elegantly derived results in a compact mathematical formalism that Einstein had derived by conceptual arguments of physical theorizing:

> Abraham had [...] achieved all essential elements of Einstein's research in the years 1907 to 1911, albeit in an entirely different way which, in addition, offered a wealth of mathematical resources for the further elaboration of a full-fledged theory of gravitation. (Renn, 2007b, p. 309)

However, Einstein soon became convinced that Abraham's approach was flawed. He continued to admire the mathematical elegance of the approach, but he rejected it on physical grounds. In Abraham's theory, the central posit of the constancy of the speed of light had to be given up because, as Einstein had shown, it varied in proportion to the gravitational potential. As a consequence, the theory was no longer globally Lorentz covariant. Einstein was not willing to accept this: he pointed out that if one uses a Lorentz covariant formulation, but global Lorentz covariance is not valid, an inner inconsistency arises. The disagreement developed into a heated exchange of papers and correspondence between Einstein and Abraham.

In the end, Einstein's view prevailed among physicists; Abraham's theory was abandoned for the reason just mentioned. Abraham nevertheless stuck by his theory and never accepted Einstein's approach. The controversy was nevertheless fruitful: the fact that Abraham's theory had failed on the whole does not imply that his use of Minkowski's formulation was doomed as well.

> Einstein [...] became gradually convinced that it was worthwhile after all to take a closer look at the utility of a modification of this formalism for his version of a gravitational field theory as well. Driven by Abraham's bold and occasionally stubborn persistence, Einstein in May 1912 thus finally recognized that a generalized line element, as suggested by Abraham's note of three months earlier, indeed represents the key to a generally relativistic gravitation theory. (Renn, 2007b, p. 316)

This change of perspective can be understood as the result of an open cycle, one round of application of mathematics that failed, but that still proved crucial for the further development of GR. Abraham's theory used the Minkowski formalism to represent SR. Einstein adopted this approach, despite its failure, because it provided him with more deductive possibilities,

[3] See Renn (2007b) for a historical account of the Einstein-Abraham controversy.

and because it suggested an alternative generalization of SR, based on the mathematical form of the line element. The reflection on the open cycle showed that, properly used, there was nothing wrong with this mathematical approach; quite to the contrary, the superior mathematical features made the Minkowski formulation attractive.

In short: what made the Minkowski approach attractive from the perspective of the IC was, first, an improvement in the deduction step of the application cycle, and, second, the anticipation that the next immersion step would be easier, if the assumed structure was formulated using the Minkowski metric.

## 8.3 Episode Two: A New Kind of Mathematics

In this section, we apply the IC to one of the most interesting episodes in the genesis of GR, the collaboration of Einstein with his "mathematician friend", Marcel Grossmann, on the so-called *Entwurf* ("outline") theory of gravitation in 1912-13; see section 5.3.2 for an introduction to the episode. The collaboration started when Einstein tried to generalize the equivalence principle, to reference frames of arbitrary acceleration, but faced mathematical difficulties. Legend has it that in a state of desperation, he turned to Marcel Grossmann, a fellow professor at the ETH Zurich, for help. Grossmann identified the Absolute Differential Calculus (ADC), an early version of tensor calculus, as a mathematical theory that could solve Einstein's problems. Together, Einstein and Grossmann reformulated gravitational theory in this new framework, and published the first tensorial formulation of GR.

In this episode, we have two goals. First, we want to reconstruct the cycle that led to the first application of the ADC to gravitational theory. Second, we want to scrutinize the evolution of a mathematical theory, the ADC, in application. These two goals complement each other, in that the first goal focuses on the application process from the perspective of one particular application, gravitational theory, while the other highlights an aspect of application that is not captured by the IC, namely the contribution of pure mathematics to its own application.

We will put to work the insights from our historical case study on Grossmann's contributions to the Entwurf theory. The episode is a particularly interesting case of the application of mathematics for the following reasons. First, the mathematics applied in the Entwurf, the ADC, were developed independently of the particular physical problem at hand, gravitational theory. This sets it apart from the application cycles of episode 1. Because of this feature, it has often been discussed as an example of the "unreasonable effectiveness" of mathematics; see Steiner (1998).

Secondly, at this stage of history, the division of labour between mathematics and physics is discernible at several levels. The protagonists of the episode, Einstein and Grossmann, both had clearly defined competences and tasks: Einstein initiated the collaboration and brought the physical motivation and knowledge to the table. Grossmann's competence was in finding a theory that solved a clearly-formulated mathematical problem. This division of labour carries over to the resulting joint publication. The Entwurf paper has two parts: Einstein was responsible for the first, physical part, while Grossmann was responsible for the second, mathematical part.

Third, the case is interesting because Einstein and Grossmann were not yet able to carry out the application process to their satisfaction, as the application cycle remained open. The characterization of the theory as an Entwurf proved to be justified, as the Entwurf field equations turned out to be wrong in the end.

We will start with a reconstruction of the episode as an application cycle, and discuss the distinctive contribution of the mathematical theories to their own application as we go.

### 8.3.1 Assumed Structure

Since the inception of the drama of GR, the assumed structure had been substantially revised; see section 5.3.2. Einstein's goal was to find relativistic field equations of gravitation, i.e. an equation of the form

$$OP(POT) = SOURCE \tag{8.1}$$

This is a so-called "frame", a template that can be instantiated differently depending on the context; see Renn and Sauer (2007, p. 127). Its instances are differential equations which determine the gravitational potential (POT) from the distribution of sources (SOURCE). In the classical case, this is the Poisson equation. At this point, Einstein had already determined two of the three components in the generalization. The SOURCE slot was filled with the energy-momentum (EM) tensor $T_{\mu\nu}$[4], and the POT slot was filled with the space-time metric $g_{\mu\nu}$. Both of these candidates resulted from several application cycles. While the EM tensor is a special-relativistic construction, the path leading to the adoption of the metric was more involved – for example, Einstein had to overcome his initial reservation vis-à-vis the Minkowski formalism.

These two components of the new relativistic field equation had become part of the assumed structure. The remaining task was to find, and examine, suitable candidates for OP, the differential operator acting on the metric, which would reduce to the Laplace operator of the Poisson equation in a suitable limit. This brings us to the immersion step.

[4]Note that in so-called vacuum solutions, the SOURCE can be the zero tensor.

### 8.3.2 Immersion Mapping

The immersion mapping connects the initial assumed structure and the mathematical domain. The goal of the immersion step was to find differential operators that could enter into the field equation. The mathematical object that had to be immersed, and that guided the search for the differential operators, was the metric. The mathematical theory had to provide differential operators acting on the $g_{\mu\nu}$ – this translates into the search for a mathematical theory that provides covariants of a homogeneous quadratic differential form.

Einstein's formulation of the task in the Entwurf paper makes it clear that the requirements on a suitable differential operator have a modeling character:

> In accordance to [the Poisson equation], one is inclined to require that [the new relativistic field equation] be of order two. However, it has to be emphasized that it proved to be impossible to find a differential equation that satisfies this requirement, is a generalization of [the Laplace operator], and is tensorial for arbitrary transformations. A priori we cannot exclude that the final, exact equations of gravitation are of order bigger than two. [...] The attempt of a discussion of such possibilities, however, would be premature in view of our present knowledge of the physical properties of the gravitational field Einstein and Grossmann (1995, p. 312).

We can extract a list of requirements, Einstein's "check list" for differential operators, from the above quote:

1. The operator should be of order two.
2. It should be invariant under transformations larger than the Lorentz group.
3. It should be a generalization of the Laplace operator.

The first is a heuristic requirement derived from the Poisson equation. The second is based on the principle of generalized relativity (see section 5.2.4 above). The third, finally, is grounded in the correspondence principle. The assumption that the operator should be of order two had no conclusive justification.

Most of the specifications for the differential operators were extracted from the existing mathematical formulation of NGT, the Poisson equation. The requirement for a wider covariance group, on the other hand, was new. Unfortunately, no historical sources concerning this issue are available: We do not know how Einstein formulated the request for candidate differential operators before he approached Grossmann for help, so it is not clear whether he thought that general covariance was required, or if he had only a vague idea that the covariance group should be larger, and Grossmann filled in the details based on the ADC. There are later recollections of the precise formulation of Einstein's question, but such later accounts always have to be taken with a pinch of salt.

### 8.3.3 The Mathematical Domain

According to the IC, the next step in the application process is the deduction step. In a certain sense, this is correct: once part of the empirical domain is represented as a mathematical structure, we use the mathematical theory to gain knowledge about this mathematical structure. However, this picture overlooks the fact that the mathematical domain, or the mathematical theory, could be a dynamical entity in itself, and could therefore make its own contributions to application. Here we will examine to what degree this is the case.

Before we do this, it may be helpful to distinguish the mathematical domain from the other steps in the application process. On the one hand, the mathematical domain is distinct from the immersion in that it is a whole mathematical framework with rules of deduction, notation, and so on. The mathematical theory needs to be able to represent the structure that is immersed into it, but it can have surplus structure. On the other hand, the mathematical domain is different from the deduction step, in that the latter is geared towards a specific goal of application. It is how the mathematical theory is put to use. The mathematical domain is a general framework which provides a space of possible structures and deductions.

In chapter 6, we discussed the mathematical part of the Entwurf, and we examined the evolution of the relevant mathematical theories prior to the Entwurf. We found that the application of the new mathematics in the Entwurf has two aspects, which we dubbed "passive" and "active" contributions.

The "passive" contributions of mathematical theories to application are those aspects that are independent of, or prior to, their application to this particular empirical domain. They concern the evolution of these mathematical theories previous to the application: What is the internal, mathematical dynamics in these mathematical theories? Did the mathematicians take any steps towards an application of their theories before Einstein and Grossmann arrived on the scene?

The "active" contributions, on the other hand, have to do with the following problems: Did Grossman adapt, change, or extend the existing mathematics in order to make it applicable? Did he make original contributions to the applied mathematics? Grossmann's own contributions can take various forms and can be of greater or minor importance, ranging from innovations regarding the notation to genuinely new mathematical results. Finally, there is always the possibility of misapplication. Mathematical theories may not be fully worked out, and there can be misunderstandings and subsequent mistakes in the application process.

Before we discuss these two kinds of contributions, it is vital to discuss which mathematical theories *are* applied in the Entwurf, and what the relative importance of these theories and traditions is.

### The Mathematical Theories in the Entwurf

We argued, in section 6.10, that Grossmann mainly drew on an algebraic-algorithmic tradition, initiated by Christoffel's work. Bianchi presented Christoffel's work in an accessible manner in his textbook on differential geometry, and Ricci & Levi-Cività expanded and developed it into a calculus in the ADC paper. In all of these contributions, the mathematical theory did not yet have an elaborated geometrical interpretation, for the most part.

The systematic significance of this choice could be the following. The ADC was formulated as a calculus with clearly defined rules of inference; it was an attractive tool, nevertheless, because it made it possible to quickly reach conclusions about the objects that could be represented in the theory. One of the central pieces of the ADC, the metric, already had geometrical interpretation. It would have been very tedious for the mathematicians to work out the geometrical significance of the other objects of the calculus, in particular the Riemann, the Ricci tensor, the manifold, or the Christoffel symbols. It was certainly more economical to do this with respect to a particular application.

### Passive Contributions: The "New Mathematics"

Let us turn to the internal dynamics of pure mathematics. We have argued, in our historical study of the mathematical traditions, that mathematicians do not only strive to solve abstract mathematical problems irrespective of applicability, but that there is a distinct tendency within pure mathematics to work towards an application of mathematical theories. These two tendencies can be nicely illustrated by comparing Christoffel's contribution, on the one hand, and Ricci & Levi-Cività's paper on the other.

Christoffel's research question is directly relevant to Grossmann: he constructed differential invariants to solve the equivalence problem of quadratic differential forms. However, his presentation is not accessible; his main goal is to solve an abstract, technical problem, not to provide an applicable tool. Ricci & Levi-Cività fix this problem by taking Christoffel's – and other mathematicians' – ideas, develop a calculus and give various examples of their application. Their paper has the goal of presenting everything in an accessible, yet general manner. Whether Ricci & Levi-Cività's efforts are entirely successful is another matter; we will return to this point below.

### Active Contributions: Adapting the New Mathematics

Grossmann's "active" contributions to tensor calculus encompass a spectrum, ranging from finding the appropriate mathematical sources and theories, to notational innovations, the reinterpretation of existing concepts and the introduction of new concepts and new mathematical results. These are all contributions that are relevant to the mathematical theory in itself, and not restricted to the application of the calculus to GR.

The most important contribution is probably the realization that the Riemann tensor is the mathematical object that allows a generally covariant approach to the field equations: all differential operators can be constructed by algebraic operations from the Riemann tensor. This approach, as we have seen in the historical part, was not yet carried out successfully in the Entwurf. To a certain degree, the same is true for the concepts of space-time metric, and of manifold. These were put to work for the first time, but the ramifications and geometrical interpretations of these concepts were not explored.

The application of the ADC to relativistic gravity left its traces in the mathematical theory. Examples are; the reinterpretation of mathematical notions; Ricci & Levi-Cività's co- and contravariant "systems" as tensors, previously a physical concept; the introduction of new concepts that are needed in application such as mixed tensors, and; new notation that is necessitated by the application, such as the distinction between co- and contravariant and mixed tensors using different kinds of letters. Many, but not all of these innovations are now standard in tensor calculus.

Grossmann also contributed new results to pure mathematics; most importantly his proofs that Beltrami parameters (differential operators), and generalizations thereof, can be given a particular form, useful in application.

### Systematic Lessons

Here are the most important systematic lessons to be learned from this case about the dynamics of the mathematical domain:

- The mathematical domain is a dynamical entity geared towards application prior to any actual application; mathematicians do not only try to solve abstract mathematical problems (think of Christoffel), but they also work towards a more accessible presentation of mathematical theories, and they strive to anticipate the needs of the prospective applicators (think of Ricci and Levi-Civita, and Bianchi), for example by developing calculi, and thereby facilitating inferences.

- As the present case illustrates, it is not a realistic expectation that a mathematical theory can just be applied without further ado. It will be necessary to modify and amend the purely mathematical theory to make it workable.

- The process of application will reflect back on the mathematics in that notions and conceptual extensions suggested by physics will become part of the mathematical theory.

- We can hypothesize that the ADC was put to work in particular because it was developed as a *calculus*, i.e. had a set of well-defined inferences. It did not yet come with any particular interpretation of the structures described by the calculus. This was an advantage, because it made the application process more economical. On the other hand, it also caused problems in the interpretation step.

### 8.3.4 Deduction Step

The deduction step, one of the focal points of the IC, takes place within the mathematical domain. In this step, we draw on a calculus, or inference rules, in order to extract mathematical knowledge about the mathematical structure in question. These inference rules can be implicit or explicit, and are ideally, but not necessarily deductive. Some deduction steps may not be fully worked out. "Material" aspects, such as notation, become important, because these can influence how easy or difficult it is to draw inferences – typically these difficulties only become apparent at this stage.

The main heuristic goal of the application cycle was to extract differential operators acting on the metric and conforming to the requirements specified in the immersion step above, and to check, in the interpretation step, whether they reduce to the Newtonian limit.

The mathematical part of the Entwurf theory contains not one, but several, application cycles, and thus several deductions. They can be found in paragraph four, the "mathematical supplement to the physical part" (see section 6.9 for an overview and discussion). The "mathematical supplement" has three parts, and accordingly three application cycles.

In the first part of the supplement, Grossmann shows that the Energy-Momentum Conservation equation is generally covariant. This is an important result, because it establishes that the conservation principle could be written in generally covariant form. Grossmann used the ADC, as well as on some of his own results, on the form of differential operators in paragraph 2. In the third part of the supplement, he provides some steps of the derivation of the "Entwurf operator". This derivation is successful, and quite elementary.

The second part of the supplement discusses the generally covariant approach to the field equations, i.e. it examines generally covariant differential operators generalizing the Laplace operator. This is the "main" application cycle, as it was the original goal to formulate a generally covariant theory of gravitation. The deduction step has two parts. First, Grossmann established that under the heuristic requirements specified in the immersion step, the Riemann tensor is *the* differential operator from which all other possible operators that could enter into the field equation can be generated. Grossmann then contracted two indices of the Riemann tensor to find a two-index tensor – the Ricci tensor – which could enter into the field equation.

Grossmann does not exploit all deductive possibilities provided by the calculus. There are other two-index tensors that can be generated from the Riemann tensor. Consulting the mathematical sources, we find that it would have been possible to find, say, the Einstein tensor, but that the deductive possibilities were not easy to grasp from the presentation in Ricci & Levi-Cività's paper, let alone from Christoffel's paper. This is a failure both of the mathematical literature, which did not make the deductive possibilities sufficiently explicit, and of Grossmann, as he did not fix the problem.

We will see below that the deduction steps in the first and third parts of the supplement are part of closed cycles. The deduction step in the second part, however, is part of an open cycle. The problem with this cycle did not become apparent in the deduction step, and only surfaced in the interpretation step.

### 8.3.5 Interpretation Step

In the interpretation step, some of the results of the deduction step are mapped back to the empirical domain, and compared either with empirical results or theoretical background knowledge. This step is not purely mathematical, as empirical and other theoretical considerations come into play – the results of the deduction are brought into contact with the assumed structure.

The interpretation step of the "main" application cycle failed: the candidate differential operator, the Ricci tensor, did not yield the Newtonian limit that Einstein expected, and thus violated the main heuristic requirement, the correspondence principle. The details of this step cannot be adequately assessed on the basis of the Entwurf paper alone; it is necessary to take into account the Zurich notebook, which documents Einsteins' struggle with the generally covariant approach to GR. We will turn to this issue in the next episode.

## 8.4 Episode Three: Not-So-Smooth Operators

In this final episode, we will put the IC to work as an account that helps us understand failures of application, the *unsuccessful* application of mathematics. The search for the field equations of GR is a story of many failed attempts at finding a suitable differential operator for the field equations. Can we discern a pattern in the problems that Einstein faced in his search? Which part of the application cycle was responsible for the problems? The IC will help us classify, and understand, Einstein's problems and mistakes.

The working hypothesis of this episode is that we can locate the problems in the components of the open cycles, the steps of the application process. However, we have to distinguish between two different kinds of reasons as to why an application cycle is open. On the one hand, an open cycle can be due to an objective mismatch between mathematics and the world, as when a mathematical theory is just unsuitable for a particular empirical problem. On the other hand, a cycle can be open due to a mistake somewhere in the cycle. To give two examples, there can be a mistake in the calculation, or some aspect of a mathematical theory can be erronenously interpreted realistically.

Many of the open cycles in the search for a differential operator fall into the category of application mistakes. As in the case of objective problems of application, the IC provides a framework for thinking systematically about application mistakes: we can locate the mistakes in the components of application cycles. Here is a short description of the four possible kinds of mistakes.

1. **Assumed Structure Mistake**: This mistake occurs when there are wrong expectations about the starting point of an application cycle, or when there are wrong expectations about the empirical target structure one expects to recover when completing a cycle.

2. **Immersion Step Mistake**: This is the mistake of taking some empirical phenomenon and choosing an unsuitable mathematical representation for it. It can happen that we have a clear mathematical counterpart for *one* empirical object, but that it is unclear as to what the appropriate representation of other aspects of reality will be. For example, knowing that the line element represents distances between space-time points does not solve the problem of how to represent space-time points.

3. **Deduction Step Mistake**: These are mistakes that occur in the deduction step, such as errors in calculations or the failure to fully exploit the deductive possibilities of a mathematical theory.

4. **Interpretation Step Mistake**: This is the mistake of interpreting part of the mathematical formalism a) realistically, if the mathematical object has properties that are purely representational, or b) not realistically, if it allows for a realistic interpretation. The most important example are coordinate systems: Einstein interpreted them realistically, but they are in fact mere tools of representation.

We will try to classify all mistakes we encounter in this episode as one of these four cases. We will pay special attention to problems with this classification scheme, as these could help us get a better picture of application, which could lead to a refined version of the IC. In particular, there could be mistakes that have to be assigned to more than one of the four cases, and there could be mistakes which are not easily attributed to any one of the four cases.

### 8.4.1 The Ricci Cycle

The first open cycle we will analyze is the one we already considered above: the application of the Ricci tensor as a differential operator for the field equation (see section 5.4.1). Now, our focus is not on the use of the ADC in application to gravitational theory, but rather on the reflection on how the open application cycle can help us understand Einstein's erroneous rejection of the Ricci tensor, and his justification for this rejection.

The reconstruction of Einstein's Zurich notebook shows that Einstein had certain expectations for how the field equations would reduce to the Newtonian limit, in particular that the weak, static limit would be spatially flat. This expectation turned out to be wrong. It was an unwarranted generalization of his previous applications of the equivalence principle, where he had explored metrics in which only the $g_{44}$ component was variable.

Which of the four kinds of mistake is this? We think this is an assumed structure mistake. Einstein had expectations for the structure that he wanted to recover at the end of the application cycle, and these expectations were unwarranted. However, there is no immersion, deduction, or interpretation mistake. All the calculations are valid, and, with certain caveats, the Ricci tensor is the right differential operator.

Einstein's reflection on the open cycle shows that he was not ready to accept this outcome on faith. Because he was convinced that his choice of assumed structure was correct, he had to locate the problem elsewhere in the cycle. The reflection convinced Einstein that the problem was due to the class of mathematical objects suggested by the ADC, the generally covariant differential operators. As he took the Ricci tensor to be the only viable candidate, a generally covariant approach was out of the question, and the covariance group had to be restricted. He thus located the mistake in the immersion step: a generally covariant operator is not the right generalization of the Laplacian operator.

However, Einstein needed a positive argument to reject all generally covariant operators. We can interpret Einstein's later formulation of the Hole Argument as an attempt to give an independent reason for this rejection of the Ricci tensor. It does not merely restate the failure to close this application cycle, but tries to explain it. We will discuss the Hole Argument further in section 8.4.3.

In a sense, then, when he failed to close the cycle, Einstein tried to establish a different equilibrium. He had good reasons to endorse the generally covariant approach provided by the ADC; most notably, generally covariant field equations promised to comply with the equivalence principle to the fullest extent. Therefore, it was necessary to find an equally good argument to reject this approach. It was not sufficient to simply give it up on the grounds that the cycle was open, as the problem may have been with the deduction step, or with the assumed structure, which would have spoken against a rejection of the generally covariant approach.

### 8.4.2 The November Cycle

The November tensor was the second differential operator that Einstein rejected in the Zurich notebook (see section 5.4.2). The rejection of the November tensor is different from that of the Ricci tensor in that, objectively speaking, it is not a suitable candidate for a generalized theory of relativity, as it presupposes a restriction of covariance to unimodular transformations. However, it is nevertheless instructive to examine Einstein's reasons for rejecting the November tensor, as they are signs of deeper problems in applying mathematics.

In retrospect, Einstein identified two problems with the November tensor. Firstly, it did not yield the correct Newtonian limit. Unfortunately, it is not clear what exactly went wrong in this step, so it is difficult to categorize the mistake. Secondly, Einstein interpreted the derivatives of the metric, and not the Christoffel symbols, as the components of the gravitational field.

There are to ways to categorize this "fateful prejudice", as Einstein called it. It could be an interpretation step mistake, the failure to recognize which part of the formalism has a realistic counterpart, or what that counterpart is. On the other hand, if the problem was in identifying the mathematical counterpart of the components of the gravitational field, then it is an immersion step mistake.

An additional difficulty that led to the second mistake was a lack of deductive possibilities. At a later stage, Einstein was able to recognize the Christoffel symbols as the correct representation, by using more powerful variational techniques. Thus, a deduction step mistake may have contributed to the "fateful prejudice".

Einstein's own diagnosis as to why he rejected the November tensor is an instance of an open cycle where different kinds of mistakes interact: a problem in the deduction step, together with the expectation of a simple result of the calculation, acted as a confirmation of a mistake Einstein made in the interpretation of the formalism.

Let us now turn to current reconstructions of the problem with the November tensor. Einstein scholars think that the deduction step mistake has a different status than the immersion or interpretation step mistake because Einstein did not finish the calculation. He thus had to ascribe the problem to his lack of mathematical ability, not to an objective problem with the theory. However, Einstein expected his theory to yield a simple solution, and his calculations may have suggested that this was not a likely outcome. Therefore, he may have interpreted the problem in the deduction step as a real, "objective" obstacle for the theory.

The majority of Einstein scholars have identified a different reason why Einstein rejected the November tensor, namely the distinction between coordinate conditions and coordinate restrictions. Einstein interpreted conditions, such as harmonic coordinates or the Hertz condition, as genuine restrictions on the covariance group of the field equations.

This can be categorized as an interpretation step mistake, the failure to recognize which parts of the formalism are to be interpreted realistically – the coordinate conditions do not have real significance in the general theory, but only in the classical limit. The mistake is not purely mathematical, as such conditions can be interpreted as coordinate conditions or as coordinate restrictions – the distinction lies in the interpretation, not in the formalism. However, some of the mathematical deductions do not make sense, if the expressions are coordinate conditions, as the calculations served to determine the covariance group of an operator under certain constraints.

John Norton has proposed an alternative account as to why the November tensor was rejected; see section 5.4.2. He thinks that Einstein was aware of the distinction between coordinate conditions and coordinate restrictions, but that he had a reason, analogous to the Hole Argument, for rejecting the November tensor. The argument presupposes that we can conceive of a coordinate condition as an independent entity, which directly represents space-time points, and from which we can remove the metric, and add a metric with different components later. This is an interpretation step mistake: Einstein failed to see that coordinates in themselves do not have a realistic interpretation. The Hole Argument only makes sense under this assumption.

The November tensor is an interesting construction from the point of view of the application of mathematics. When he explored the Ricci tensor, Einstein convinced himself that a generally covariant differential operator was not an option, and that the requirement of general covariance had to be weakened, which meant that substantial modifications of the candidates provided by the ADC were necessary. However, the fact that Einstein constructed the November tensor based on the Ricci tensor suggests that Einstein was not yet ready to give up on the mathematical possibilities provided by the ADC. The November tensor was still part of the mathematical strategy. Thus, even though Einstein was unable to close the cycle, i.e. to successfully apply the ADC in the first attempt, he did not simply reject the mathematical framework, but rather continued to exploit it.

### 8.4.3 Remarks on the Hole Argument

John Norton has conjectured that the Hole Argument may have been relevant for Einstein's rejection of the November tensor. In the philosophy of physics, the Hole Argument has risen to prominence in the debate on space-time substantivalism: the position that space-time is, to some extent, independent of the events taking place in it.[5] In modern terms, the Hole Argument rules out one particular form of substantivalism, so-called *manifold substantivalism*. This is the position that the four-dimensional space-time manifold itself represents space-time events.

Norton's discussion shows that the Hole Argument can be formulated independently of manifolds – all we need are coordinate functions with a realistic interpretation, i.e. that coordinate functions directly refer to space-time points. Here we will have a look at the argument from a historical point of view. In particular, we will examine how the mathematical literature, prior to the application of the ADC to GR, may have influenced the argument, and whether the perspective of the IC is fruitful in this context.

In section 6.5, we discussed the history of the manifold concept in mathematics, prior to its application in GR, and Grossmann's manifold concept in the Entwurf. We saw that the concept was not as clear-cut as the modern notion, and we distinguished several interpretations. The most important difference between the modern and the historical concept is that from the modern perspective, we first define manifolds as sets of points with local differentiable structure, and assign the metric to the manifold afterwards, whereas from a historical perspective, the metric and the manifold were, at least according to some interpretations, not distinct entities: the manifold is just the geometrical entity described by the line element.

Ricci & Levi-Cività define manifolds as follows in the introduction of their paper:

> A manifold $V_n$ is defined intrinsically in its metrical properties by $n$ independent variables and by a whole class of quadratic differential forms of these variables, any two of which can be transformed into each other by a point transformation. As a consequence, the $V_n$ is invariant under all transformations of its coordinates.[6]

[5]See section 5.4.2 for a historical exposition of the argument and a short discussion of Norton's conjecture. Norton (2011) is an accessible discussion of space-time substantivalism in the context of the Hole Argument.

[6][U]ne variété $V_n$ est définie intrinsèquement dans ses propriétés métriques par $n$ variables indépendantes et par toute une classe de formes quadratiques des différentielles de ces variables, dont deux quelconques sont transformables l'une en l'autre par une transformation ponctuelle. – Par conséquence une $V_n$ reste invariée vis-à-vis de toute transformation de ses coordonnées. (Ricci and Levi-Civita, 1901, p. 128)

A manifold is not defined independently of, or prior to, the metric, but by the variables, and by *classes* of the metric, at the same time. Thus the Hole Argument does not apply to this notion of manifold: defining the manifold by classes of metrics blocks the possibility of thinking of the metric as an entity that is independent of the points of the manifold, and interpreting two metrics in the same equivalence class as different entities. This rules out the construction of the Hole Argument, which relies on the distinctness of $g_{\mu\nu}$ and $g'_{\mu\nu}$, even if the two metrics are related by a coordinate transformation.

The situation is more complicated in Grossmann's case. On some interpretations, the Hole Argument does not stand up for Grossmann's manifold concept. In particular, if manifolds are close to, or even identified with, the line element – an invariant quantity – different metrics in the same equivalence class are nothing but notational variants.

Thus, historically speaking, manifolds are not responsible for the confusion leading to the Hole Argument. Rather, the mistake can be traced back to an interpretation of coordinate functions as direct representations of space-time points, a sort of "coordinate substantivalism". But how did this problem come about? How could this confusion arise, given that the mathematical literature discussed the ADC as a calculus that allowed physical theories to be formulated independently of coordinate systems?

One part of the problem can be explained by the focus of the mathematical literature. Christoffel's and Ricci & Levi-Cività's work describes how distances in space-time can be represented in a coordinate-independent manner, namely as the invariant line element. This implies that the components of the metric do not have a realistic interpretation, in that they depend on the coordinate functions we use – they are not invariants. However, the mathematical literature does not answer the question as to how to represent space-time points, or events. The formalism was not designed for this particular application, and the applicators were on their own when it came to finding the correlate of space-time points in the ADC; this made the misconceptions possible.

A second aspect of the problem could stem from the kind of application mistake that Einstein and Grossmann committed. It is easier to commit a mistake in the interpretation step when we try to find empirical correlates for parts, or aspects, of the formalism. It is at least *prima facie* plausible to ascribe some property in the world to the coordinate system. On the other hand, if the question was what the correct mathematical correlate of space-time points are, then it is a) harder to disregard the invariant-theoretic perspective of the formalism, and b) easier to take prior immersions of other mathematical objects into account, and model the later immersion on these prior examples – in the present case, on the immersion of distances in space-time as the line element.

Finally, a big part of the problem was that the relevant application cycles are not independent. Once we have a mathematical representation for distances, the question as to how to immerse space-time points cannot be answered independently of the representation of distances, and it is rather the earlier step that determines the answer: we cannot speak about space-time points independently of the metric.

In the end, one cannot help but wonder at the irony that, if Einstein had accepted the manifold concept as proposed by Ricci & Levi-Cività, he could have avoided the confusion arising from the Hole Argument, and he could have defended a sort of *manifold* substantivalism* – manifold* being the historical, "invariant" notion as opposed to the modern manifold concept.

### 8.4.4 Systematic Observations

Here are some of the systematic lessons learned from an analysis of the open application cycles in the search for a differential operator.

First, the conceptual tools provided by the IC allow us to locate application mistakes in a descriptively adequate manner. There are only a few cases in which we cannot locate the mistakes in the framework. The analysis in terms of the IC is also philosophically insightful, in that it makes it easier to further analyze, what could be called, the dynamics of application, that is, the influence of misconceptions on the further course of the application process.

Second, the distinction between an immersion step mistake and an interpretation step mistake seems to be instructive. These mistakes have a different logic in that one is a misconception about the realistic interpretation of mathematics, while the other is a misconception about the right representation of a phenomenon or aspect of the world.

Third, not all the components of the cycle are on a par. Problems with either immersion or interpretation can lead to the outright rejection of a mathematical theory for a particular application. A problem with the deduction step, on the other hand, is more likely attributable to the applicator's inexperience, or an incomplete understanding of the possibilities provided by a mathematical framework. In the present case, it is interesting that Einstein attributed an "objective" significance to a problem in the deduction step, his "fateful prejudice", in that, by doing so, he took the simplicity of a mathematical expression to be a sign of its verisimilitude.

Fourth, the fact that one particular aspect of the world is successfully mirrored in mathematics, i.e. that there is a closed cycle, does not guarantee the successful application in other cases – quite to the contrary, it can complicate matters. For example, the ADC provided a framework for capturing space-time distances as an invariant quantity: the line element. However, it was not clear how space-time points should be mirrored in that framework. To find the right representation for space-time points was all the more difficult because it is not possible to characterize them independently of the metric. Thus, the subsequent application cycles are not independent;

latter cycles have to be compatible with former cycles, which can make the application task increasingly hard.

## 8.5 Summary

Let us sum up our most important findings.

In episode one, we examined the initial assumed structure, the empirical phenomena which are the starting point of the application process. Not all relevant gravitational phenomena are part of the assumed structure – the anomalous precession of Mercury's perihelion did not play a role in the formulation of GR, but rather served as a test for the completed theory.

We found that the assumed structure was given by the empirical content of SR and NGT. However, it was an important part of the discovery of GR to identify the right mathematical formulation of these two theories, namely the Minkowski formalism, which facilitated inferences and suggested the invariant line element as a central object that had to be generalized, and the Poisson equation, which was helpful for a field-theoretic approach, and not only served to capture empirical content, but also played a theoretical role in the new theory of GR.

We suggest that settling for these two formulations was already part of the application process, and that the importance of the mathematical form of the predecessor theories in the discovery and justification of GR is paramount.

In episode two, we reconstructed the first application of the ADC to gravitation based on the IC framework. The application cycle, leading to the application of the ADC, was triggered by the search for a differential operator acting on the metric with a larger covariance group. Einstein and Grossmann did not succeed in closing this cycle, as they found that the candidate differential operator, the Ricci tensor, does not reduce to the expected Newtonian limit. They did, however, close two other cycles, one of them yielding the Entwurf operator with restricted covariance.

Based on the historical study of the mathematical part of the Entwurf paper, we conjecture that the mathematical domain itself has to be treated as a separate aspect of application. We found that some mathematicians work on abstract mathematical puzzles with negligible ties to direct application, as witnessed by Christoffel's work, while others, such as Ricci & Levi-Cività, strive to make mathematical theories applicable, facilitating inferences, and even suggesting potential domains of application. This contribution of pure mathematics to its own application is not captured by the IC.

We also found that Grossmann himself put a considerable effort into the application of the ADC to gravitation, contributing to many aspects of the mathematical theory. The application of mathematics it is not a matter of plug-and-play, it is an active process that transforms the mathematics through the application process.

Finally, we noted that the first application of the ADC did not go smoothly. For one, Grossmann failed to realize the full deductive potential of the ADC, when he only derived the Ricci tensor as a viable differential operator, and neglected other possible candidates.

In episode three, we used the IC as conceptual framework for analyzing application mistakes in Einstein's search for a differential operator for the field equations. We reconstructed his rejection of the Ricci tensor and the November tensor, documented in the Zurich notebook, as open cycles. We found that the IC provides useful categories for the analysis of application mistakes, such as the distinction between immersion step mistakes and interpretation step mistakes. Further systematic observations can be found in section 8.4.4.

Finally, we briefly commented on the Hole Argument, from the perspective of the application of mathematics. The Hole Argument can be interpreted as an attempt to give a positive reason for the failed attempt to close the Ricci cycle. We then found that, historically, the notion of manifold is not responsible for the confusion arising from the argument, because at least some historical conceptions of the notion would have blocked the argument.

All in all, the extended IC provides a useful framework for the analysis of the application process, and allows a nuanced reconstruction of the historical episodes, if one keeps in mind that it does not capture the internal application dynamics of mathematics.

# Conclusion

The goal of this thesis was to discuss the following questions: What is the role of mathematics in application to the world? and, why is mathematics useful in solving empirical problems? In the conclusion, I would like to gather some systematic lessons learned from the various case studies, and point out avenues for future research.

## Metaphysical Issues

The problem of applicability is related to some intricate metaphysical questions, which can be divided into three groups: the metaphysics of mathematics, the metaphysics of the empirical domain, and the nature of the relation between these two domains.

I critically discussed one particular account of the metaphysics of mathematics, *ante rem* structuralism, in chapter 1. The alternative I sketched in this context – the representation of mathematical structures in terms of isomorphism types – is not convincing from a metaphysical point of view; isomorphism types are not the answer to the question as to what are mathematical structures. However, I am rather sceptical that we will ever be able to provide a concluding answer to this question.

Instead, I propose to pay more attention to the question as to what particular mathematical theories and structures are applied in particular contexts, and to think about the philosophical ramifications of these choices. First, in the case study of GR, it proved fruitful to examine the status, and traditions behind, tensor calculus – we discerned an invariant-theoretic and a geometrical perspective, and argued that, at least in the beginning, the former prevailed, with all its advantages and drawbacks. Second, in the Königsberg case, I argued that, although, strictly speaking, we do not need graph theory to solve the Königsberg problem, graph theory is still important at the level of pure mathematics, e.g. for reducing complexity.

As to the metaphysics of the empirical domain, I argued, both in the Königsberg case and the Volterra case, that we can give relevant parts of the mathematical structures a causal interpretation. Keeping this in mind prevents us from taking the mathematical structures themselves metaphysically seriously. The causal interpretation can inform our view as to what kind of explanation we are dealing with; in particular, none of the cases we saw are "purely mathematical" explanations of physical phenomena. I will further comment on explanations below. In the honeycomb case, the causal interpretation of the structure is even more important, as the construction mechanism of the physical structure can speak in favor, or against, the applicability of mathematical results.

In the case study on GR, we did not enter into the classical debates on the metaphysics of space-time theories, but rather limited ourselves to a discussion of the Hole Argument, as well as a discussion about which parts of the mathematical formalism of GR *can* be given a realistic, geometrical interpretation. Here we got the impression that many of the interpretational problems of the genesis of GR could have been settled through a careful handling of the available mathematical theory, the ADC, which is, after all, a theory of mathematical objects that are independent of coordinate systems. However, we will have to pursue these issues further.

Finally, there is the question as to how the relation between mathematical and empirical domains is constituted. Critics of the Inferential Conception point out that if the correspondence between mathematical and empirical domains is spelled out in terms of mappings, the result risks being circular, as mappings are themselves mathematical entities. The assumed structure problem raises similar concerns.

One of the common themes of all the cases we review is that the relation between mathematics and the world is not one of simple structural correspondence. First, I argued that, in the Königsberg case, the representation is, at least partially, a matter of pragmatics. Second, idealizations are paramount in all cases. This shows that a direct correspondence between mathematics and empirical structure is out of the question. Thus, the objection that the IC is circular loses traction. However, it also shows that we should not interpret the mappings of the IC as establishing a direct structural correspondence.

## Explanation

All of the smaller case studies of applicability – the Königsberg case, the bee's honeycomb, and the predator-prey model – are examples of explanations where mathematics plays a central role. What lessons can we learn from these cases about explanations, be they scientific or within mathematics?

We can wholeheartedly reject the notion that there are mathematical explanations, in the sense that we can explain empirical phenomena exclusively on the basis of mathematics. In all cases we examined, we saw that a mere match between an empirical and a mathematical structure was not sufficient to explain the empirical structure; think of the explanation of the hexagonal openings based on the HC, or the reproduction of periodically fluctuating populations by the predator-prey model.

The bridge principle between mathematical structures and the world is often highly non-trivial and in need of independent justification. I argued that, in the Königsberg case, it has a pragmatic justification. In the case of the bee's honeycomb, I pointed out that, even if we succeed in physically reproducing the actual honeycomb as a kind of foam, we should suspend judgement as to whether this constitutes an explanation, because we do not know whether the underlying mechanism producing the structure, the liquid equilibrium, is the actual construction process. Finally, the reproduction of the Third Law, as a mathematical analogue of the population shift, is not sufficient for accepting the model as an explanation of this phenomenon. Only if we can reproduce the Third Law in more realistic models, should we accept this account.[7]

In all cases we examined, it proved to be fruitful, or, at least, possible, to find a causal interpretation for the mathematical structure or model. Because of this, I also rejected the idea that scientific explanations using mathematics are non-causal. I think that those maintaining such views simply project purely mathematical explanations onto the world.

In the Königsberg chapter, I proposed that mathematics can contribute to scientific explanations at two levels: the relation between mathematics and the world, on the one hand, and the explanation of purely mathematical facts, on the other. This suggests that we should accept explanations in pure mathematics, which can, but need not, take the form of proofs. I further proposed several factors that potentially contribute to purely mathematical explanations; the two most promising being reduction of (computational) complexity, and elimination of irrelevant information. This proposal is essentially a refined version of Mark Steiner's transmission view.

## Idealization

One of the common themes of the cases we considered is the issue of idealization. Volterra explicitly discusses, and tries to justify, the idealizations that go into the predator-prey model. However, he is ultimately not successful, as the model is now commonly taken to be flawed. I argued that the same is true for Lyon and Colyvan's application of the HC to the bee's honeycomb, and also for Fejes Tóth's mathematical approach to the honeycomb.

I also argued that the introduction of idealizations is not *per se* responsible for the fact that these models are flawed. Euler's explanation is also based on idealizing assumptions, but we nevertheless accept it as a good ex-

[7] One caveat is in order: I did not examine a *fundamental*, physical theory, such as QM, with respect to explanation. Maybe the situation is different in these cases, as we may not have access to the underlying mechanisms of fundamental theories.

planation. The same is true for most mathematical models – all models are wrong, but some are useful, as the statistician George E. P. Box quipped. Why are certain idealizations harmless, while others can render a model virtually useless for certain purposes? And: what's mathematics got to do with it?

It is very difficult to give systematic reasons as to why certain idealizations are adequate for certain purposes, while others are not, and the inductive basis of the cases we considered is certainly too thin. However, one discernible pattern is that the *mathematics-driven* idealizations, i.e. ideas motivated by, say, mathematical simplicity, or tractability, or by the desire to find an application of a particular mathematical result, were particularly problematic.

Mathematics-driven idealizations are not necessarily problematic. It is possible to find a good justification for the choice of a particularly simple mathematical model *ex post facto*. However, such a justification always has to be supplied later, as simplicity need not track truth – this is in stark contrast to *empirically-driven* idealizations, which already come with a (good or bad) justification, by their very nature.

Thus, it could be speculated that there exists asymmetry in the justification of these two kinds of idealizations – empirically-driven idealizations come with a built-in justification, which can be good or bad, while mathematics-driven idealizations are always in need of later justification – and that this asymmetry is responsible for the fact that the latter are particularly problematic: when we use mathematics-driven idealizations, we have to get lucky, while we have already weeded out the worst empirically-driven idealizations.

We did not say much about the idealizations of the "modeling exercise" of GR. This does not mean that no idealizations went into the construction of the Einstein field equations. For example, the equations are only unique under the assumption that they should be of order two, and linear in the second derivatives; see Steiner (1998, pp. 94) for more on this issue. While the analogy with the Poisson equation goes some way towards justifying this choice, it is also based on considerations of mathematical simplicity – at least, this is Einstein's perspective. Thus, we are, again, dealing with a mathematics-driven idealization. It would certainly be worthwhile to further explore the justification of these assumptions.

## The IC: Prospects and Problems

In this thesis, we examined accounts of the applicability of mathematics; in particular, we tried to apply the IC to GR. How successful are these attempts? Should we try to refine the IC, and similar accounts, or should they be given up?

On the positive side, the IC has proven to be a useful tool in the description, and analysis, of the historical interplay between mathematics and physics in the case study of GR. What is more, this account seems to capture, at least some practitioners' conception, of the application of mathematics; see Volterra's reflections on the role of mathematics.

I think that the usefulness of the IC lies its simplicity, and in that it allows us to decompose a big problem into smaller components ("divide and conquer"). The account does not presuppose much. All we have to assume is that there exists a clear separation between the empirical and the mathematical domain, and that there is some kind of correspondence between the two. The IC breaks up the application process into components, i.e. manageable steps, which can be analyzed separately, and then be reassembled.

However, the IC is a complete account of applicability. As we pointed out in the case study on GR, there are aspects of applicability about which the account is silent; most importantly, it neglects the internal dynamics of mathematics. Also, the problems of conceiving of the relation between the two domains as of mere structure-preservation were pointed out throughout the thesis; other philosophical problems are open as well. Thus, if we continue using the IC as a framework for the process of applying mathematics, we should always keep in mind that it does not faithfully capture all relevant aspects of applicability. It should be used as a conceptual and heuristic tool, not as the final word on applicability.

## History and Discovery

What are the advantages of examining the historical genesis and the construction process of mathematical theories and models?

Many of the systematic issues we examined benefitted from the historical approach. To take two obvious examples; the examination of Euler's original paper uncovered several approaches to the Königsberg problem, which were fruitfully applied in the discussion on scientific and mathematical explanations; and the discussion of Volterra's justification of idealizations, which revealed the primary motivations for idealizations, suggesting that some of them were mathematics-driven.

Of course, a historical approach is indispensable for questions of theory dynamics. However, examining the actual construction process of scientific theories substantially changes our perspective on the relation between mathematics and the world. It becomes apparent that the process of application is one of mutual adaptation: empirical questions and desiderata can lead to substantial changes, and extensions, of mathematical theories; and the mathematical expression of empirical phenomena can lead to substantial conceptual changes in the way we see the world, and even help in the discovery of new phenomena. All these aspects come to the fore once we conceive of the application of mathematics as of a historical process.

Finally, the historical approach also reveals problems of applicability, notably mistakes. These are usually not visible, if only the result of the application process is taken into account – think of Einstein's three-year struggle with his field equation. However, a systematic understanding of the pitfalls of the application process yields a better understanding of applicability *tout court*. We are at the very beginning of this journey.

## Unreasonable Effectiveness?

What are the consequences of our case studies for the thesis of the unreasonable effectiveness of mathematics? Of course, we cannot reject the thesis on the basis of just a few examples. However, we certainly can discern whether the cases speak in favor, or against, the thesis.

All the cases we saw, except for the Königsberg case, are not pure success stories, but rather involve a considerable struggle with the mathematics; in some of the cases, the application even fails altogether. It is questionable to call the mathematics effective, at all, in these cases.

The case of GR is particularly noteworthy. The application of the ADC to gravitational theory is certainly effective, but not unreasonably so. One of the presuppositions of the thesis of the unreasonable effectiveness is that the methods of mathematics are mainly driven by aesthetic considerations, and independently of empirical considerations. In the case of the ADC, this is not correct. Even in Christoffel's paper, which could be considered to be farthest from real-world applications, the connection to a real, geometrical problem is quite clear – the equivalence problem of homogeneous quadratic differential forms, a formal problem, translates into the problem of which metrics are related by coordinate transformations. In the paper by Ricci & Levi-Cività, the methodological gap is even smaller. All in all, the assumption that mathematical research is mainly driven by aesthetics just seems to be wrong.

Then, as we pointed out repeatedly, the mathematics was not "plug and play", but rather there was a considerable effort by Einstein and Grossmann to extend the existing mathematics. The suggestion, that mathematics can be applied without further ado, is plainly wrong.

This is not to say that there are no gaps in our understanding of the application of mathematics in the genesis of GR. I already mentioned the formal restrictions on the form of the final field equations in the context of idealizations above. Some of the properties of the field equations can be derived form the Poisson equation, and there are physical reasons for doing so, namely the correspondence principle. However, this does not completely determine the field equations – at this point, mathematical simplicity comes into play. I think we should not interpret this as a case of unreasonable effectiveness, but rather as an idealization that is still in need of proper justification.

## Concluding Remarks

I believe that the applicability of mathematics is an important philosophical issue, which deserves more attention from philosophers and scientists, and I am convinced that a unified perspective on the problems of applicability could be beneficial for many debates in the philosophy of science and mathematics.

Clearly, all the cases I discussed in this thesis should be further explored; in particular, the case study on GR, which is a very rich source that we have only begun to understand. I would certainly like to continue working on this case.

There are two areas that I consider to be particularly interesting from the perspective of applicability, but which I did no touch on in the present thesis. The first is quantum mechanics, and the second is the use of mathematical models in economics, and in finance in particular. If I have one regret, it is that I did not have time to take a closer look at these two important examples of the application of mathematics.

# Bibliography


Baker, A. 2005. Are there Genuine Mathematical Explanations of Physical Phenomena? *Mind* 114(454): 223–38.

———. 2009. Mathematical Explanation in Science. *British Journal for the Philosophy of Science* 60(3): 611–633.

———. 2012. Science-Driven Mathematical Explanation. *Mind* 121(482): 243–67.

Baker, A. and M. Colyvan. 2011. Indexing and Mathematical Explanation. *Philosophia Mathematica* 19(3): 323–34.

Batterman, R. W. 2002. *The Devil in the Details. Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford, New York: Oxford University Press.

———. 2010. On the Explanatory Role of Mathematics in Empirical Science. *British Journal for the Philosophy of Science* 61(1): 1–25.

Bauer, D. and K. Bienefeld. 2013. Hexagonal Comb Cells of Honeybees are Not Produced Via a Liquid Equilibrium Process. *Naturwissenschaften* 100: 45–49.

Benacerraf, P. 1965. What Numbers Could not Be. In P. Benacerraf and H. Putnam, eds., *Philosophy of Mathematics*. Cambridge: Cambridge University Press, 2nd ed., pp. 272–94.

Bianchi, L. and M. Lukat. 1899. *Vorlesungen über Differentialgeometrie*. Leipzig: Teubner, 1st ed.

Brauer, F. and C. Castillo-Chávez. 2001. *Mathematical Models in Population Biology and Epidemiology*, vol. 40 of *Texts in Applied Mathematics*. New York: Springer.

Braun, M. 1993. *Differential Equations and Their Applications*, vol. 11 of *Texts in Applied Mathematics*. New York, Berlin, Heidelberg: Springer-Verlag, 4th ed.

Bueno, O. and M. Colyvan. 2011. An Inferential Conception of the Application of Mathematics. *Nous* 45(2): 345–74.

Burgess, J. P. 1999. Book Review: Stewart Shapiro, Philosophy of Mathematics. *Notre Dame Journal of Formal Logic* 40(2): 283–91.

Cantor, G. 1895. Beiträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen* 46(4): 481–512.

Cartwright, N. 1983. *How the Laws of Physics Lie.* Oxford, New York: Oxford University Press.

Christoffel, E. B. 1869. Ueber die Transformation der homogenen Differentialausdrücke zweiten Grades. *Journal für reine und angewandte Mathematik* 70: 46–70.

Colyvan, M. 2001. *The Indispensability of Mathematics.* Oxford, New York: Oxford University Press.

———. 2009. Mathematics and the World. In A. D. Irvine, ed., *Handbook of the Philosophy of Science: Philosophy of Mathematics.* North Holland: Elsevier, pp. 651–702.

———. 2011. Indispensability Arguments in the Philosophy of Mathematics. Http://plato.stanford.edu/entries/mathphil-indis/.

———. 2012. *An Introduction to the Philosophy of Mathematics.* Cambridge Introductions to Philosophy. Cambridge: Cambridge University Press.

———. 2013. The Undeniable Effectiveness of Mathematics in the Special Sciences. In M. C. Galavotti, S. Hartmann, M. Weber, W. Gonzalez, D. Dieks, and T. Uebel, eds., *New Directions in the Philosophy of Science.* Springer.

Dell'Aglio, L. 1996. On the Genesis of the Concept of Covariant Differentiation. *Revue d'histoire des mathématiques* 2: 215–64.

Diestel, R. 2006. *Graph Theory.* Graduate Texts in Mathematics. Berlin, Heidelberg: Springer.

do Carmo, M. P. 1993. *Differentialgeometrie von Kurven und Flächen.* Braunschweig: Vieweg & Sohn, 3rd ed.

Ehlers, J. 1981. *E. B. Christoffel: The Influence of His Work on Mathematics and the Physical Sciences*, chap. Christoffel's Work on the Equivalence Problem for Riemannian Spaces and Its Importance for Modern Field Theories of Physics. Basel, Boston, Stuttgart: Birkhäuser, pp. 526–42.

Einstein, A. and M. Grossmann. 1995. Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation. In Klein et al. (1995), pp. 302–43.

Euler, L. 1956. The Seven Bridges of Königsberg. In J. R. Newman, ed., *The World of Mathematics*, vol. 1. Simon and Schuster, pp. 573–580.

Fejes Tóth, L. 1964. What the Bees Know and What they do not Know. *Bulletin AMS* 70: 468–81.

Field, H. 1980. *Science Without Numbers. A Defence of Nominalism.* Princeton: Princeton University Press.

French, S. 2000. The Reasonable Effectiveness of Mathematics: Partial Structures and the Application of Group Theory to Physics. *Synthese* 125: 103–20.

French, S. and D. Krause. 2006. *Identity in Physics.* Oxford: Clarendon Press.

Frigg, R. and S. Hartmann. 2012. Models in Science. Http://plato.stanford.edu/entries/models-science/.

Hafner, J. and P. Mancosu. 2005. The Varieties of Mathematical Explanation. In Mancosu et al. (2005), pp. 215–50.

———. 2008. Beyond Unification. In Mancosu (2008), pp. 151–78.

Hales, T. C. 2000. Cannonballs and Honeycombs. *Notices of the AMS* 47(4): 440–9.

———. 2001. The Honeycomb Conjecture. *Discrete and Computational Geometry* 25: 1–22.

Halmos, P. 1960. *Naive Set Theory.* Princeton, New Jersey: D. Van Nostrand Company, Inc.

Hellman, G. 2001. Three Varieties of Mathematical Structuralism. *Philosophia Mathematica* 9(2): 184–211.

———. 2005. Structuralism. In S. Shapiro, ed., *The Oxford Handbook of Philosophy of Mathematics and Logic*, chap. 17. Oxford, New York: Oxford University Press, pp. 536–62.

Hepburn, J. R., T. Muerrle, and S. E. Radloff. 2007. The cell bases of honeybee combs. *Apidologie* 38(3): 268–71.

Hermann, R. 1975. *Ricci and Levi-Civita's Tensor Analysis Paper*, vol. II of *Lie Groups: History, Frontiers and Applications.* Brookline, Massachusetts: Math Sci Press.

Hopkins, B. and R. J. Wilson. 2004. The Truth about Königsberg. *The College Mathematics Journal* 35(3): 198–207.

Jakob, C. 2007. *Wissenschaftstheoretische Grundlagen sozial- und geschichtswissenschaftlicher Erklärungen*. Ph.D. thesis, Universität Bern.

Janssen, M., J. D. Norton, J. Renn, T. Sauer, and J. Stachel, eds. 2007a. *The Genesis of General Relativity*, vol. 1. Einstein's Zurich Notebook: Introduction and Source. Dordrecht: Springer.

———. 2007b. *The Genesis of General Relativity*, vol. 2. Einstein's Zurich Notebook: Commentary and Essays. Dordrecht: Springer.

Janssen, M. and J. Renn. 2007. *Untying the Knot: How Einstein Found His Way Back to Field Equations Discarded in the Zurich Notebook*. Vol. 2. Einstein's Zurich Notebook: Commentary and Essays of Janssen et al. (2007b), pp. 839–925.

Jech, T. 1997. *Set Theory*. Berlin, Heidelberg, New York: Springer, 2nd ed.

Keränen, J. 2001. The Identity Problem for Realist Structuralism. *Philosophia Mathematica* 9(3): 308–30.

Ketland, J. 2006. Structuralism and the identity of indiscernibles. *Analysis* 66(4): 303–15.

———. 2011. Identity and Indiscernibility. *Review of Symbolic Logic* 4(2).

Kingsland, S. E. 1985. *Modeling Nature. Episodes in the History of Population Ecology*. Chicago and London: The University of Chicago Press.

Klarreich, E. G. 2000. Foams and Honeycombs. *American Scientist* 88(2): 152–61.

Klein, M. J., A. J. Kox, J. Renn, and R. Schulmann, eds. 1995. *The Collected Papers of Albert Einstein*, vol. 4: The Swiss Years: Writings, 1912-1914. Princeton, New Jersey: Princeton University Press.

Kottler, F. 1912. *Ueber die Raumzeitlinien der Minkowski'schen Welt*, vol. 121 of *Mathematisch-naturwissenschaftliche Klasse. Abteilung IIa. Sitzungsberichte*. Wien: Kaiserliche Akademie der Wissenschaften.

Kruja, E., J. Marks, A. Blair, and R. Waters. 2002. A Short Note on the History of Graph Drawing. In *Lecture Notes in Computer Science*, vol. 2265. Berlin, Heidelberg: Springer-Verlag, pp. 272 – 86.

Ladyman, J., O. Linnebo, and R. Pettigrew. 2010. Identity and discernibility in philosophy of logic. *The Review of Symbolic Logic* 5(1): 162–86.

Ladyman, J. and D. Ross. 2007. *Every Thing Must Go.* Oxford, New York: Oxford University Press.

Lange, M. 2013. What Makes a Scientific Explanation Distinctively Mathematical? *British Journal for the Philosophy of Science* 64(3): 485–511.

Lehtinen, A., J. Kuorikoski, and P. Ylikoski, eds. 2012. *Economics For Real. Uskali Mäki and the Place of Truth in Economics.* New York: Routledge.

Leitgeb, H. and J. Ladyman. 2008. Criteria of Identity and Structuralist Ontology. *Philosophia Mathematica* 16(3): 388–96.

Lyon, A. 2012. Mathematical Explanations Of Empirical Facts, And Mathematical Realism. *Australasian Journal of Philosophy* 90(3): 559–78.

Lyon, A. and M. Colyvan. 2008. The Explanatory Power of Phase Spaces. *Philosophia Mathematica* 16(2): 227–243.

Maddy, P. 1997. *Naturalism in Mathematics.* Oxford: Clarendon Press.

———. 2007. *Second Philosophy. A Naturalistic Method.* Oxford: Oxford University Press.

Mancosu, P., ed. 2008. *The Philosophy of Mathematical Practice.* Oxford, New York: Oxford University Press.

Mancosu, P. 2011. Explanation in Mathematics. Http://plato.stanford.edu/entries/mathematics-explanation/.

Mancosu, P., K. F. Jorgensen, and S. A. Pedersen, eds. 2005. *Visualization, Explanation and Reasoning Styles in Mathematics*, vol. 327 of *Synthese Library.* Dordrecht: Springer.

May, R. M. 2001. *Stability and Complexity in Model Ecosystems.* Princeton Landmarks in Biology. Princeton, Oxford: Princeton University Press.

Morgan, F. 1988. *Geometric Measure Theory. A Beginner's Guide.* Burlington, San Diego, London: Academic Press, 3rd ed.

Moschovakis, Y. 2006. *Notes on Set Theory.* Undergraduate Texts in Mathematics. New York: Springer, 2nd ed.

Murray, J. D. 1993. *Mathematical Biology*, vol. 19 of *Biomathematics Texts.* Berlin, Heidelberg: Springer, 2nd ed.

Norton, J. D. 1984. How Einstein Found His Field Equations: 1912-1915. *Historical Studies in the Physical Sciences* 14: 253–315.

———. 1992. *The Physical Content of General Covariance*, vol. III of *Einstein Studies.* Boston: Birkhäuser, pp. 281 – 315.

———. 2000. ‘Nature is the Realization of the Simplest Conceivable Mathematical Ideas’: Einstein and the Canon of Mathematical Simplicity. *Studies in History and Philosophy of Modern Physics* 31(2): 135–70.

———. 2005. A Conjecture on Einstein, the Independent Reality of Spacetime Coordinate Systems an the Disaster of 1913. In A. J. Kox and J. Eisenstaedt, eds., *The Universe of General Relativity*, vol. 11 of *Einstein Studies*. Basel, Boston, Berlin: Birkhäuser, pp. 67 – 102.

———. 2007. *What Was Einstein's "Fateful Prejudice"?* Vol. 2. Einstein's Zurich Notebook: Commentary and Essays of Janssen et al. (2007b), pp. 715–83.

———. 2011. The Hole Argument. Http://plato.stanford.edu/entries/spacetime-holearg/.

Pais, A. 1982. *Subtle is the Lord ...: The Science and the Life of Albert Einstein*. Oxford, New York: Oxford University Press.

Pincock, C. 2004. A New Perspective on the Problem of Applying Mathematics. *Philosophia Mathematica* 12(3): 135–61.

———. 2005. Overextending Partial Structures: Idealization and Abstraction. *Philosophy of Science* 72: 1248–59.

———. 2007. A Role for Mathematics in the Physical Sciences. *Nous* 41(2): 253–75.

———. 2012. *Mathematics and Scientific Representation*. Oxford, New York: Oxford University Press.

Pirk, C. W. W., H. R. Hepburn, and S. E. Radloff. 2004. Honeybee combs: construction through a liquid equilibrium process? *Naturwissenschaften* 91(7): 350–3.

Potter, M. 2004. *Set Theory and its Philosophy*. Oxford, New York: Oxford University Press.

Räz, T. 2013a. Comment on "The Undeniable Effectiveness of Mathematics in the Special Sciences". In M. C. Galavotti, S. Hartmann, M. Weber, W. Gonzalez, D. Dieks, and T. Uebel, eds., *New Directions in the Philosophy of Science*. Springer.

———. 2013b. On the Application of the Honeycomb Conjecture to the Bee's Honeycomb. *Philosophia Mathematica* .

Reich, K. 1989. Das Eindringen des Vektorkalküls in die Differentialgeometrie. *Archive for History of Exact Sciences* 40: 275–303.

———. 1994. *Die Entwicklung des Tensorkalküls: Vom absoluten Differentialkalkül zur Relativitätstheorie*. Historical Studies. Basel, Boston, Berlin: Birkhäuser.

Renn, J. 2007a. *Classical Physics in Disarray*. Vol. 1. Einstein's Zurich Notebook: Introduction and Source of Janssen et al. (2007a), pp. 21–80.

———. 2007b. *The Summit Almost Scaled: Max Abraham as a Pioneer of a Relativistic Theory of Gravitation*. Vol. 3. Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy of Renn and Schemmel (2007), pp. 305–30.

Renn, J. and T. Sauer. 2007. *Pathways out of Classical Physics. Einstein's Double Strategy in his Search for the Gravitational Field Equation*. Vol. 1. Einstein's Zurich Notebook: Introduction and Source of Janssen et al. (2007a), pp. 113–312.

Renn, J. and M. Schemmel, eds. 2007. *The Genesis of General Relativity*, vol. 3. Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy. Dordrecht: Springer.

Resnik, M. D. and D. Kushner. 1987. Explanation, Independence and Realism in Mathematics. *The British Journal for the Philosophy of Science* 38(2): 141–58.

Ricci, G. 1904. Direzione e Invarianti Principali in Una Varietà Qualunque. *Atti del Reale Instituto Veneto* 63(2): 1233–39.

Ricci, M. and T. Levi-Civita. 1901. Méthodes de calcul différentiel absolu et leurs applications. *Mathematische Annalen* 54: 125–201.

Riemann, B. 1876a. Commentatio mathematica, qua respondere tentatur quaestioni ab IIIma Academia Parisieni propositae. In H. Weber, ed., *Gesammelte Mathematische Werke und wissenschaftlicher Nachlass*. Teubner, 1st ed., pp. 370–83.

———. 1876b. *Gesammelte Mathematische Werke und wissenschaftlicher Nachlass*. Leipzig: Teubner, 1st ed.

———. 1876c. Ueber die Hypothesen, welche der Geometrie zu Grunde liegen. In H. Weber, ed., *Gesammelte Mathematische Werke und wissenschaftlicher Nachlass*. Leipzig: Teubner, 1st ed., pp. 254–69.

———. 1892. *Gesammelte Mathematische Werke und wissenschaftlicher Nachlass*. Leipzig: Teubner, 2nd ed.

Rindler, W. 2006. *Relativity. Special, General, and Cosmological*. Oxford, New York: Oxford University Press, 2nd ed.

Rotman, J. J. 1995. *An Introduction to the Theory of Groups.* New York: Springer, 4th ed.

Roughgarden, J. 1997. *Primer of Ecological Theory.* Upper Saddle River: Prentice-Hall.

Saatsi, J. 2011. The Enhanced Indispensability Argument: Representational versus Explanatory Role of Mathematics in Science. *British Journal for the Philosophy of Science* 62: 143–54.

Sachs, H., M. Stiebitz, and R. J. Wilson. 1988. An Historical Note: Euler's Königsberg Letters. *Journal of Graph Theory* 12(1): 133–9.

Sandborg, D. 1998. Mathematical Explanation and the Theory of Why-Questions. *British Journal for the Philosophy of Science* 49(4): 609–24.

Sauer, T. 2013. Marcel Grossmann and His Contribution to the General Theory of Relativity. In *forthcoming.*

Scholl, R. and T. Räz. 2013. Modeling causal structures. *European Journal for Philosophy of Science* 3(1): 115–32.

Scholz, E. 1980. *Geschichte des Mannigfaltigkeitsbegriffs von Riemann bis Poincaré.* Boston, Basel, Stuttgart: Birkhäuser.

Schutz, B. 2009. *A First Course in General Relativity.* Cambridge: Cambridge University Press, 2nd ed.

Shapiro, S. 1997. *Philosophy of Mathematics: Structure and Ontology.* Oxford, New York: Oxford University Press.

———. 2008. Identity, Indiscernibility, and *ante rem* Structuralism: The Tale of *i* and -*i*. *Philosophia Mathematica* 16(3): 285–309.

Simon, B. 1996. *Representations of Finite and Compact Groups*, vol. 10 of *Graduate Studies in Mathematics.* American Mathematical Society Publications.

Stachel, J. 1980. Einstein and the Rigidly Rotating Disk. In A. Held, ed., *General Relativity and Gravitation. One Hundred Years After the Birth of Albert Einstein*, vol. 1. New York, London: Plenum Press, pp. 1–15.

———. 2007. *The First Two Acts.* Vol. 1. Einstein's Zurich Notebook: Introduction and Source of Janssen et al. (2007a), pp. 81–112.

Steiner, M. 1978a. Mathematical Explanation. *Philosophical Studies* 34(2): 135–51.

———. 1978b. Mathematics, Explanation, And Scientific Knowledge. *Nous* 12(1): 17–28.

———. 1998. *The Applicability of Mathematics as a Philosophical Problem*. Cambridge Mass., London: Harvard University Press.

———. 2005. Mathematics – Application and Applicability. In S. Shapiro, ed., *The Oxford Handbook of Philosophy of Mathematics and Logic*, chap. 20. Oxford: Oxford University Press, pp. 625–50.

Szpiro, G. G. 2003. *Kepler's Conjecture*. Hoboken, New Jersey: Wiley.

Tallant, J. 2013. Optimus prime: paraphrasing prime number talk. *Synthese* 190(12): 2065–83.

van Fraassen, B. 1980. *The Scientific Image*, chap. The Pragmatics of Explanation. Oxford: Clarendon Press.

Volterra, V. 1926. Fluctuations in the Abundance of a Species Considered Mathematically. *Nature* 118: 558–60.

———. 1928. Variations and fluctuations of the number of individuals in animal species living together. *J. Cons. int. Explor. Mer* 3(1): 3–51.

Volterra, V. and U. D'Ancona. 1935. *Les associations biologiques au point de vue mathématique*. Paris: Hermann.

Weaire, D. and S. Hutzler. 1999. *The Physics of Foams*. Oxford: Clarendon Press.

Weaire, D. and R. Phelan. 1994. Optimal design of honeycombs. *Nature* 367(13): 123.

Weisberg, M. and K. Reisman. 2008. The Robust Volterra Principle. *Philosophy of Science* 75(1): 106–31.

Wigner, E. P. 1995. The Unresonable Effectiveness of Mathematics in the Natural Sciences. In *Philosophical Reflections and Syntheses*. Springer, pp. 534–49.

Wilholt, T. 2004. *Zahl und Wirklichkeit: Eine philosophische Untersuchung über die Anwendbarkeit der Mathematik*. Paderborn: Mentis.

Wilson, R. J. 1986. An Eulerian Trail Through Königsberg. *Journal of Graph Theory* 10(3): 265–75.

Zund, J. D. 1983. Some Comments on Riemann's Contributions to Differential Geometry. *Historia Mathematica* 10: 84–9.