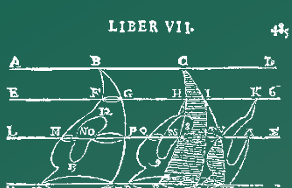
Proceedings of the Canadian Society for History
and Philosophy of Mathematics
La Société Canadienne d'Histoire
et de Philosophie des Mathématiques

Maria Zack
Dirk Schlimm Editors

Research in History and Philosophy of Mathematics

The CSHPM 2016 Annual Meeting in Calgary, Alberta



Birkhäuser

Birkhäuser

Proceedings of the Canadian Society for History and Philosophy of Mathematics La Société Canadienne d'Histoire et de Philosophie des Mathématiques

Series Editors
Maria Zack, MICS Department, Point Loma Nazaren Univ, San Diego, CA, USA
Dirk Schlimm, McGill University, Montreal, QC, Canada

More information about this series at http://www.springer.com/series/13877

Maria Zack • Dirk Schlimm
Editors

Research in History and Philosophy of Mathematics

The CSHPM 2016 Annual Meeting in Calgary, Alberta

Birkhäuser

Editors
Maria Zack
Mathematical, Information and Computer Sciences
Point Loma Nazarene University
San Diego, CA, USA

Dirk Schlimm
Department of Philosophy
McGill University
Montreal, QC, Canada

ISSN 2366-3308 ISSN 2366-3316 (electronic)
Proceedings of the Canadian Society for History and Philosophy of Mathematics
La Société Canadienne d'Histoire et de Philosophie des Mathématiques
ISBN 978-3-319-64092-1 ISBN 978-3-319-64551-3 (eBook)
DOI 10.1007/978-3-319-64551-3

Library of Congress Control Number: 2017956715

Mathematics Subject Classification: 01-06, 01A72, 01A99

Printed on acid-free paper

This book is published under the trade name Birkhäuser, www.birkhauser-science.com
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume contains fourteen papers that were presented at the 2016 Annual Meeting of the Canadian Society for History and Philosophy of Mathematics (CSHPM), which was held at the University of Calgary in Calgary, Alberta, in May of 2016. The papers are arranged in roughly chronological order and contain an interesting variety of modern scholarship in both the history and philosophy of mathematics.

The life and work of the logician and philosopher of mathematics Aldo Antonelli (1962–2015) was honored at this conference and the first four papers in this book are part of that remembrance. All of these scholarly works are philosophical in nature. The volume begins with a discussion of the nineteenth-century polymath Bernard Bolzano. Bolzano is widely seen as a figure who showed that even in Kant's time, there were philosophical and mathematical reasons to think that an intuition-based philosophy of mathematics was untenable. In "Bolzano Against Kant's Pure Intuition," Paul McEldowney identifies and clarifies two of Bolzano's longstanding objections to Kant's philosophy of mathematics. In "Poincaré on the Foundation of Geometry in the Understanding," Jeremy Shipley discusses Poincaré's core commitment in the philosophy of geometry: the view that geometry is the study of groups of operations.

In "Natural Deduction for Diagonal Operators," Fabio Lampert presents a sound and complete Fitch-style natural deduction system for an **S5** modal logic containing an actuality operator, a diagonal necessity operator, and a diagonal possibility operator. Finally in "What is a Symbol?" Valerie Allen argues that the symbolic notation under development in the sixteenth and seventeenth centuries has connections with rhetorical and poetic aesthetics and through etymological analysis of the polysemy of its terminology she considers the philological orientation of early modern algebra.

The history of mathematics papers begin with Robert Bradley's "Polar Ordinates in Bernoulli and L'Hôpital." In this paper, Bradley discusses early schemes for polar ordinates which are described in the Marquis de l'Hôpital's *Analyse des infiniment petits* (1696), and based on the lessons given to the Marquis by Johann Bernoulli. In the seventeenth century, the Polish Jesuit and polymath Adam Adamandy Kochański studied magic squares, and in 1686 he published a paper in *Acta Eruditorum* titled "Considerationes quaedam circa Quadrata et Cubos Magicos." In that paper he proposed a novel type of magic square, where in every row, column, and diagonal, if the entries are sorted in decreasing order, the difference between the sum of entries with odd indices and those with even indices is constant. Henryk Fukś discusses these unusual magic squares in "Magic Squares of Subtraction of Adam Adamandy Kochański."

In "Euler's E228: Primality Testing and Factoring via Sums of Squares," Fred Rickey looks at a clever method that Euler devised for determining if a number is a sum of two squares. Euler applied this method to 1,000,009, and in less than a page found that there are two ways to express this as a sum of squares. Euler was the mentor of Anders Lexell and in "'A Most Elegant Property': On the Early History of Lexell's Theorem" Eisso J. Atzema discusses both Euler and Lexell's proof of a result in spherical geometry that is now known as Lexell's Theorem.

The term "Playfair's Axiom" is a mainstay of school geometry textbooks as well as one of the few things many people know about the eighteenth-century mathematician John Playfair. In "The Misnamings of Playfair's Axiom," Amy Ackerberg-Hastings discusses the interesting history of this axiom and what it reveals about the history of nineteenth-century mathematics education. In "Napier, Torporley, Menelaus, and Ptolemy: Delambre and De Morgan's Observations on Seventeenth-Century Restructuring of Spherical Trigonometry," Joel Silverberg analyzes De Morgan's criticism of Napier's and Torporley's efforts to organize spherical trigonometry and uses the analysis to shed light on the challenges to a historian of mathematics of one era attempting to understand the thought process of mathematicians living in earlier times.

The final papers in the volume all look at aspects of twentieth-century mathematics. In "The Reception of American Mathematics Education in Soviet Pedagogical Journals of the 1960s and 1970s," Mariya Boyko discusses reforms in Soviet mathematical instruction. She examines the reception of American mathematics education changes known as "new math" in Soviet pedagogical journals of the 1960s and 1970s and provides a comparison of the Russian experience with what took place in the West.

In "Mathematics in Library Subject Classification Systems" Craig Fraser discusses the modern classification of mathematical subjects within the larger framework of library classification in the early twentieth century. This paper explores different views during this period concerning the position of mathematics in the overall scheme of knowledge, the scope of mathematics, and the internal organization of the different parts of mathematics with a focus on the Library of Congress classification system in its various iterations from 1905 to the present. Finally, in "The Convolution as a Mathematical Object" Roger Godard considers the modern mathematical properties of the convolution in the context of the work of Volterra, Lebesgue, Doetsch, and Schwartz and looks at specific applications of the convolution in the theory of probability, interpolation, and smoothing.

The book concludes with the delightful “Sundials – Une Promenade Parisienne.” In this brief paper George Heine takes the reader on a walk through Paris to look at a variety of interesting examples of sundials and provides a method for estimating the accuracy of historic calendrical sundials.

This collection of papers contains several gems from the history and philosophy of mathematics which will be enjoyed by a wide mathematical audience. This collection was a pleasure to assemble and contains something of interest for everyone.

San Diego, CA, USA Maria Zack
Montreal, QC, Canada Dirk Schlimm

Editorial Board

The editors wish to thank the following people who served on the editorial board for this volume:

Contents

Contributors

Amy Ackerberg-Hastings Independent Scholar, Adelphi, MD

Valerie Allen John Jay College CUNY, New York, NY, USA

Eisso J. Atzema University of Maine, Orono, ME, USA

Paul Anh McEldowney Department of Philosophy, University of Notre Dame, Notre Dame, IN, USA

Mariya Boyko University of Toronto, Toronto, ON, Canada

Robert E. Bradley Department of Mathematics and Computer Science, Adelphi University, Garden City, NY, USA

Craig Fraser University of Toronto, Toronto, ON, Canada

V. Frederick Rickey Professor Emeritus of Mathematics, United States Military Academy, West Point, NY, USA

Henryk Fukś Department of Mathematics and Statistics, Brock University, St. Catharines, ON, Canada

Roger Godard Department of Mathematics and Computer Science, Royal Military College of Canada, Kingston, ON, Canada

George Heine Independent Scholar, Pueblo, CO

Fabio Lampert University of California, Davis, CA, USA

Jeremy Shipley Volunteer State Community College, Gallatin, TN, USA

Joel S. Silverberg Professor Emeritus, Roger Williams University, Providence, RI, USA

Bolzano Against Kant's Pure Intuition

Paul Anh McEldowney

Abstract The 19th-century polymath Bernard Bolzano is widely regarded as a figure who decisively showed that even in Kant's time, there were philosophical and mathematical reasons to think that an intuition-based philosophy of mathematics was untenable. This paper identifies and clarifies two of Bolzano's longstanding objections to Kant's philosophy of mathematics. Once this is done, I defend Kant from these objections. To conclude, I discuss a deeper disagreement between Bolzano and Kant's philosophical approaches. Namely, Bolzano rejected that a "critique of reason" (i.e., an examination of the principles and limits of human cognition) was necessary in order to secure and explain reliable forms of inference and claims to a priori knowledge.

Keywords Bernard Bolzano • Immanuel Kant • Intuition

1 Introduction

In *The Semantic Tradition from Kant to Carnap*, Alberto Coffa praises the 19th-century mathematician and philosopher Bernard Bolzano for founding a distinctively anti-Kantian "semantic" tradition, which Coffa defines as follows:

> The semantic tradition may be defined by its problem, its enemy, its goal, and its strategy. Its problem was the a priori; its enemy, Kant's pure intuition; its purpose, to develop a conception of the a priori on which pure intuition played no role; its strategy, to base that theory on a development of semantics. (Coffa 1991, p. 22)

In terms that parallel Coffa's appraisal, Bolzano saw himself as taking up the task of showing how the "path ... which the *Critique* pursues ... is not at all the right one" (Příhonský 1850, p. 146). According to Bolzano, in order to provide a philosophically satisfying account of the nature and possibility of a priori knowledge, one should begin with a careful examination into "nature of

P. Anh McEldowney (✉)
Department of Philosophy, University of Notre Dame, Notre Dame, IN, USA
e-mail: pmceldow@gmail.com

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*,
Proceedings of the Canadian Society for History and Philosophy of Mathematics
La Société Canadienne d'Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_1

the truths in themselves" (Ibid., p. 146) rather than an analysis of our cognitive capacities, as Kant mistakenly does. In this paper, I seek to identify what Bolzano found to be problematic about Kant's intuition-based approach to mathematical knowledge. This will not only put us in a better position to determine the resilience of Kant's views against Bolzano's criticisms, but it will allow us to appreciate the philosophical problems that Bolzano's own views seek to overcome.

The effect of Kant's critical philosophy on Bolzano's thought cannot be underestimated, as Bolzano held a longstanding philosophical engagement with Kant's critical philosophy and Kantian philosophers such as Schultz and Kiesewetter. In fact, Bolzano's influential 1810 treatise, *Beträge zu einer begründeteren Darstellung der Mathematik*[1] (*BD*), includes an appendix dedicated to arguing why Kant's notion of *pure intuition* was both incoherent and unable to ground mathematical knowledge. Bolzano continues his quarrel with Kant in his 1837 magnum opus, the *Wissenschaftslehre*[2] (*WL*). In addition to presenting global criticisms of Kant's critical philosophy, Bolzano's *WL* seeks to provide viable alternatives to various Kantian positions. Following the publication of *WL*, Bolzano's philosophical engagement with Kant is reinvigorated by the writings of Bolzano's pupil František Příhonský. With the consultation and approval of Bolzano, Příhonský, in 1850, published *Neuer Anti-Kant*[3] (*NAK*), a systematic commentary on Kant's first *Critique* from the point of view of Bolzano's *WL*. As the title suggests, Příhonský's *NAK* was far from a positive assessment of Kant's first *Critique*.[4]

Despite Bolzano's disagreements with Kant, it is important to take note of some of the illuminating similarities between the respective overarching philosophical projects of these two thinkers. Just as Kant thought that philosophy was stuck in a "battlefield of [...] endless controversies" (Kant 1781, Aviii), Bolzano thought errors plaguing the sciences were responsible for why "in mathematics one also finds long disputes over certain subjects, e.g., the measure of vis viva, the ratio of force to resistance with wedges, the distribution of pressure over several points, etc., which are still unresolved" (Bolzano 1837, §315). While Kant sought to resolve disputes in metaphysics by distinguishing the method of philosophy from that of mathematics, Bolzano sought to resolve disputes in the sciences by instigating a program of foundational reform that critiqued the correctness of and standards for the definitions, axioms, and proofs used in the sciences, especially those in mathematics. In order to preserve the status of the sciences as paradigms of human knowledge, Bolzano thought their presentation had to reflect objective ground-consequence relations between something like mind-independent propositional contents.

[1] "Contributions to a better-grounded Presentation of Mathematics."

[2] "Theory of Science."

[3] "New Anti-Kant."

[4] More specifically, *NAK* was instigated by Bolzano in 1837, the same year *WL* was published. Příhonský was responsible for writing and publishing *NAK*, and Bolzano was actively involved in its development, providing Příhonský with input and eventually, his approval of the finished product. Thus, there is good reason to think that *NAK* reflects Bolzano's own views.

Like other critics of Kant at the time, Bolzano took issue with Kant's claim that constructions in pure intuition could ground mathematical knowledge.[5] This paper focuses on two closely related objections of Bolzano's, both of which reflect critical themes recurring throughout his works:

CAT: By its nature, intuition cannot account for the generality of mathematical cognition or the necessity of mathematical inference.[6]

DIS: Even if intuition can help discover that a mathematical proposition is true, intuition plays a dispensable role in securing mathematical knowledge and inference.

For convenience, I'll call the first objection, the *categorical objection*, or CAT, and the second, the *dispensability objection*, or DIS. The purpose of this paper is to clarify these two objections from the point of view of Bolzano's corpus. In order to properly assess the force of these objections, it will be important to distinguish the various roles that Kant claims intuition to play in mathematical cognition and inference. Once these objections are clarified, my plan is to defend Kant against them. Thus, while Coffa (1991), Rusnock (1999), Lapointe (2011), and others champion Bolzano as decisively showing that Kant's views on mathematics were untenable even in Kant's time, I argue that neither the categorical nor the indispensability objection do the job. To conclude, I briefly discuss what I take to be fundamentally driving Bolzano's dissatisfaction with Kant's view: Bolzano rejected that a "critique of reason" was necessary to explain and secure reliable forms of synthetic inference and claims to a priori knowledge.

2 Kant's Theory of Cognition

Let's begin with a brief overview of the basics behind Kant's theory of cognition as explicated in the first *Critique*. For Kant, cognition (i.e., "Erkenntnis") is a technical term used to refer to one's awareness of an object or objects as having determinate

[5]The critics I have in mind are those who ascribe to what might be suitably called a "Wolffian" or "Leibnizian" philosophy of mathematics according to which a mathematical proof of a proposition *P* (which is necessary and sufficient to furnish mathematical knowledge of *P*) proceeds and "depends on the principle of contradiction and in general on the doctrine of syllogisms" (Lambert 1761, §92.14). See Heis (2014) for more on this debate between the Wolffian and the Kantian.

Secondly, it is important to point out that, according to Bolzano, there is a distinction between the actual reasons why a proposition *P* is true and what merely convinces us that *P* is true. The former tracks objective ground-consequence relations mentioned in the previous paragraph. These relations hold regardless of whether there are any minds or epistemic agents.

Bolzano was not the first to make this distinction. In fact, the distinction between the metaphysical ordering of truths and the doxastic order by which one comes to believe a truth can be traced back to Aristotle (1993) in Book I of the *Posterior Analytics*.

[6]In this paper, I understand a proposition *P* to be "general" or "universal" if *P* applies to a class of objects such as all triangles or all natural numbers and not just particular instances. I take "necessity" to mean something similar except that it is a property that applies primarily to arguments and inferences.

properties. As such, cognition is a type of epistemic achievement that results from the interplay between two distinct faculties of representation, namely sensibility and understanding. Through sensibility we are affected passively by the world, receiving what Kant calls "intuitions". Intuitions are particular representations that refer immediately to an object. Contrasting with the passive faculty of sensibility, understanding is an active faculty responsible for unifying representations into "concepts", "thoughts", and "judgments". Concepts are general (i.e., repeatedly applicable) representations that relate to an object immediately through marks.[7]

Since mathematical cognitions are arguably non-empirical (i.e., they are not about anything, nor can they be justified by anything one encounters in experience), the story becomes slightly more complicated. As Kant argues, instances of the concepts of mathematics are not given in experience. Rather, to grasp a mathematical concept for Kant is *to be able to construct it* in pure intuition according to a general procedure given by the concept's construction schema, which is broadly understood to be something like its genetic definition.[8] Kant argues in the *Transcendental Aesthetic* that space and time are a priori intuitions. As such, they constitute the forms of sensibility; i.e., they are precisely those invariant and necessary features of experience and how objects are given to us.

To give an example of what Kant has in mind when he speaks of construction in pure intuition, Kant would argue that in order to grasp the concept <triangle>, one would have to construct a triangle in pure intuition according to its construction schema. From there, one can reason with one's construction and prove, for instance, the Angle Sum Theorem, i.e., the sum of the angles of a triangle is two right angles. Similarly, an intuition of a homogeneous unit constructed in the pure intuition of time underwrites the concept of quantity as such needed to reason in arithmetic or algebra. From a Kantian perspective, when one begins a proof with the phrase "let n be an arbitrary natural number," one is reasoning with a constructed homogeneous unit.

I argue that constructions in pure intuition are meant to play at least two important roles in Kant's account of mathematical cognition:

[7] To give a mundane mathematical example, the concept <triangle> (I will use angle brackets to denote concepts) has among its marks the concept <three-sided>.

[8] Classically understood, genetic definitions are definitions that reveal the "genesis" or "cause" of the thing being defined. What this means is often spelled out in terms of a finite procedure that specifies how to construct the object being defined presumably from a given or privileged set of entities. For instance, a genetic definition for a triangle provides a procedure of constructing a triangle from more basic kinds of objects (e.g., points and lines) by certain rules and methods of (perhaps, straightedge and compass) construction. No claim here is made as to whether for Kant, genetic definitions have to be *unique*. Friedman (2010) construes schemata as *functions* of a certain type. They are functions whose inputs (e.g., three arbitrary lines) are a given set of entities and whose output is a representation (e.g., a triangle constructed from those three lines). Precisifying "schemata" in terms of genetic definitions or functions (in Friedman's sense) works equally well for the purposes of the paper.

AX: Constructions are needed in order to have cognition of axioms as self-evident truths that are general in nature.

INF: Constructions are needed to justifiably make necessary and general inferences that go beyond what can be inferred by means of concept-containment relations (i.e., conceptual analysis) or traditional logic (i.e., an Aristotelian term logic with additional principles regarding disjunctive and hypothetical judgments).

Before proceeding to clarify what these two roles consist in, it is important to note that the essence of Bolzano's categorical and dispensability objection is to undercut AX and INF. The categorical objection doubts whether intuition can play either of these roles, while the latter maintains that intuition—even if it has any non-trivial epistemic role at all—is entirely indispensable in acquiring mathematical knowledge.

According to Kant, axioms are *synthetic* truths that are immediately evident and unprovable.[9] As synthetic, axioms involve concepts that do not contain one another.[10] Famously, Kant argues that the axioms of Euclidean geometry are genuine paradigm cases of axioms according to his understanding of the word. Thus, in particular, the axiom asserting that "there exists a line between any two points" is *immediately known* once one constructs two points in pure intuition.

Since axioms involve concepts that do not bear the proper requisite concept-containment relations, Kant argues that one cannot arrive at cognition of axioms through conceptual analysis. Thus, conceptual analysis alone is unable to distinguish a genuine axiom from grammatically sound yet arbitrarily conjoined concepts. Since genuine cognition arises either out of conceptual analysis or with the help of intuition, Kant concludes that intuition is, in fact, *needed* to secure knowledge of axioms (i.e., AX).[11] And so Kant's pure intuition is what separates a genuine axiom from arbitrarily conjoined concepts. As a consequence of Kant's view, the cognition of the Euclidean axiom asserting that "between two points there exists a line" results immediately from constructing two points in pure intuition.

Kant contrasts axioms with judgments such as "the sum of the angles of a triangle is equal to two right angles," which are inferred from premises. According to Kant, the difference between the two is straightforward. Cognition that is inferred requires an inference and a set of premises from which the cognition is inferred. On the other hand, immediate cognition does not involve premises from which an inference is made.

[9]See, for example, Kant (1781, A303) and (Ibid., A732/B760) for clear articulations of Kant's conception of axioms, especially as they compare with knowledge that is *inferred*.

[10]Kant formulates the distinction between analytic and synthetic judgments in different possibly inequivalent ways. Nonetheless, in this paper, the distinction will be understood in terms of concept-containment relations. As already touched on, this distinction hinges on a certain view of concepts, namely that concepts have *marks*, which are just further concepts, and that concepts admit of decomposition in terms of these marks. This view of concepts was somewhat standard and shared among thinkers such as Lambert and even Bolzano.

[11]That intuition *can* furnish knowledge of axioms is taken to be evident for Kant in light of the epistemic legitimacy of Euclidean diagrammatic geometry. See Friedman (2010) for more on the relationship between Kant and Euclidean geometry.

Kant argues that constructions are not only necessary for explaining our cognition of axioms, i.e., he argues for AX, but he argues that they are needed to draw reliable inferences that go beyond the rules of general logic, i.e., he argues for INF.

According to Kant, general logic concerns the "logical form in the relation of cognitions to one another" (Kant 1781, B76–79). As such, general logic consists in the rules that hold of any subject matter regardless of whether there are any epistemic agents at all. Based on this characterization of general logic, Kant concludes that the principles of logic are exhausted by traditional logic. Regardless of what one takes as constituting the principles of logic, what matters for Kant is that they concern the mere form of thoughts.

So in particular, Kant argues that it is not possible for us to provide a proof the Angle Sum Theorem that proceeds solely by means of general logic and conceptual analysis. One needs to begin with an a priori construction of a triangle, and perform auxiliary constructions until one arrives at the theorem under question.

To summarize, Kant's philosophy of mathematics is primarily concerned with mathematical cognition as an epistemically special mental state that arises from the interplay between the faculties of sensibility and understanding. Logical and conceptual resources (i.e., general logic and concept-containment relations) are unable to account for (1) our knowledge of mathematical axioms as *immediate* and *synthetic*; or for (2) our knowledge of inferred extra-formal mathematical knowledge. Since logical or conceptual means won't do the trick, a priori constructions, according to Kant, account for and are needed to account for (1) and (2).

3 The Categorical Objection

At this point, one may already anticipate Bolzano's objections. Let's first focus on the categorical objection, which is best expressed in the following passage from Bolzano's *BD*:

> How do we come, from the intuition of that single object, to the feeling that what we observe in it also belongs to every other one? Through that which is single and individual in this object, or through that which is general? Obviously only through the latter, i.e., through the concept, not through the intuition. (Bolzano 1810, A. 7)

According to Bolzano, Kant's philosophy of mathematics falls short since Kant has yet to provide a compelling account of how intuitions, as essentially singular, can underwrite our knowledge of mathematical truths and inferences as being in any way necessary. According to Bolzano, one can only be aware of the necessity and generality of mathematical truths if one *proceeds from concepts* rather than intuitions. It seems that intuitions *by their very nature* are unable to show that a mathematical truth *has* to be true or that an inference is *valid* since intuitions are merely instantiations of concepts, and inferring from particulars to general statements is in general unreliable. This is the categorical objection in a nutshell.

In order to defend Kant from the categorical objection, one must show how particular intuitions can reliably guide an epistemic agent to cognition of general mathematical truths. As the following passage illustrates, Kant's solution is to appeal to the generality of the procedure or schema by which a concept is constructed in pure intuition:

> No image at all would ever be adequate to the concept of a triangle in general. For it would never attain the universality of the concept, which makes it hold for all triangles, whether right-angled, acute-angled, and so on, but would always be limited to only a part of this sphere. The schema of the triangle can never exist anywhere but in thought, and it signifies a rule of synthesis of the imagination with respect to pure figures in space. (Kant 1781, A140/B180)

According to Kant, since constructions proceed through schemata, whatever is known from using a construction does not depend on any of the particular determinations and magnitudes that end up being constructed in pure intuition. Thus, as Kant notes, through schemata one is able to "consider the universal in the particular" (Kant 1781, A714/B742).

Bolzano was well aware of Kant's appeal to schemata as a possible reply to the categorical objection. However, Bolzano remained unsatisfied:

> If the schema is nothing but an idea of the way in which a circle is generated, i.e., nothing but the well-known definition of a circle, which is called a genetic definition, namely, the concept of a line which is described by a point that moves in a plane in such a way that it always maintains the same distance from a given point. Thus if one recognizes the truth of a synthetic proposition through the consideration of the schema of its subject-concept, one recognizes this truth from the consideration of mere concepts. (Bolzano 1837, §305)

Ironically, an appeal to schemata seems to support Bolzano's claim that only through the use of concepts can one acquire knowledge of mathematical truths as universal in character; an inference from a genetic definition is just an inference from concepts.

In a sense, Kant *agrees* with Bolzano. The generality of mathematical cognition and the necessity of an inference in mathematics is due to the general nature of concepts and their schemata. However, synthetic mathematical cognition requires that the mathematical concepts involved be constructed in pure intuition. While one may have a rule to draw triangles, one needs to actually carry out the construction in pure intuition in order to cognize mathematical axioms and make inferences that go beyond concept containment relationships and traditional logic. As Kant notes, one can "reflect on a [mathematical] concept as long as he wants, yet he will never produce anything new" (Kant 1781, A715/B744). One must construct the concept as "through a chain of inferences that is always guided by intuition, he arrives at a fully illuminating and *at the same time* general solution to the question" (Kant 1781, A717/B745, my emphasis). To succinctly make the point, while concepts and schemata are *necessary* for mathematical cognition, they are not sufficient.

To recap, throughout his writings, Bolzano argues that intuition is not the type of thing that could ground mathematical cognition, immediate or inferred. I hope to have shown that the particularity of intuitions does not preclude them from enjoying a necessary role in mathematical cognition.

4 The Dispensability Objection

While Bolzano's categorical objection questions whether intuitions are even the right kind of thing that can underwrite mathematical cognition, the dispensability objection attacks Kant's claim that intuitions are *necessary* for mathematical cognition, immediate or inferred. Bolzano argues that intuition (a priori or otherwise) is largely inessential to mathematical knowledge.

Bolzano argues for the dispensability objection in at least two ways. First, Bolzano provides a number of counterexamples found in mathematics where intuition seems to not play a dispensable role. Second, Bolzano provides a competing and equally plausible account of mathematical knowledge to Kant's whereby intuitions play but an entirely dispensable role. In what follows, I discuss each argument independently.

4.1 Bolzano's Counterexamples

In *BD*, Bolzano provides a series of examples where intuition appears to play an inessential role in securing mathematical knowledge. For instance, Bolzano argues that as long as one assumes the associativity of addition on the natural numbers, the pure intuition of time is not needed to explain one's knowledge of arithmetical sums. Due to associativity, "one looks at the set of terms, not their order" (Bolzano 1810, A.8). One can then compute "7 + 2" by defining the natural numbers and addition recursively, i.e., by assuming that an undefined primitive term "1" denotes a number and what results from repeated applications of the successor function to "1" also denotes a number. Thus, one can compute "7 + 2" in the following way without the aid of intuition:

$$
\begin{aligned}
7 + 2 &= 7 + (1 + 1) && \text{(By definition)} \\
&= (7 + 1) + 1 && \text{(By associativity)} \\
&= 8 + 1 && \text{(By definition)} \\
&= 9 && \text{(By definition)} \\
& && \text{QED}
\end{aligned}
$$

It is not difficult to defend Kant from this purported counterexample. One can of course define and stipulate whatever one likes. Kant would argue that one can only know the correctness of definitions and stipulated rules of inference (the latter of which are either axioms or proven from axioms) if one verifies them in pure intuition. According to Kant, because arithmetic sums like "7 + 1 = 8" cannot be known from pure conceptual analysis, help must be sought in pure intuition. Without such verification, Bolzano's purported logical proof that "7 + 2 = 9" fails to exhibit the synthetic nature of arithmetical cognition that Bolzano, like Kant, attributes to computing sums.[12]

Furthermore, given that Bolzano in *BD* adopts Kant's formulation of the distinction between the analytic and synthetic, Bolzano would be rationally bound to agree with Kant that the associativity of addition on the natural numbers is a *synthetic truth*, the knowledge of which remains in need of explanation (Bolzano 1810, §17).[13]

Before proceeding, it is important to clear up a possible misunderstanding on Bolzano's end. Regardless of whether associativity is analytic or synthetic, one should note that the pure intuition of time does not ground arithmetic by virtue of the fact that we do computations in time. We certainly draw syllogistic inferences in time, and yet, such inferences are not grounded in the pure intuition of time. The pure intuition of time is the material with which one constructs homogeneous units needed to perform computations.[14] Thus, if the associativity of addition is not analytic for Kant, then our cognition of the associativity of sum is only possible if we construct numerical units in pure intuition.

4.1.1 A Counterexample from Analytic Geometry

In a similar spirit to the above purported counterexample, Příhonský in *NAK* argues that the rise of analytic geometry has shown that constructions in pure intuition are inessential even in the context of geometry:

> Rather [than relying on intuition, it] is possible to deduce all truths of geometry from the correct definition of space, without ever allowing even once a conclusion that has no other ground for justification in favour of it besides what visual inspection teaches. The well-known analytical geometry provides us with examples of this procedure that show at least how many geometrical truths which visual inspection teaches can also be brought out without any appeal to it, through mere inferences. (Příhonský 1850, p. 116)

According to Příhonský, one can solve problems in geometry by positing systems of equations and a coordinate system without any appeal to visualization or

[12]It might be helpful to recall Russell's criticism of "the method postulating" by which one postulates mathematical axioms and definitions without verifying their truth and consistency. According to Russell, without doing the hard philosophical and mathematical work of verifying the axioms and definitions that one has laid down, the method of postulation amounts to nothing more than "theft over honest toil" (Russell 1919, p. 71). Thus, Bolzano cannot appeal to stipulated definitions and axioms in order to undermine Kant's philosophy of mathematics since the validity of such definitions and axioms are the very thing intuitions are supposed to secure, according to Kant. I thank one of the anonymous referees for pointing this out to me.

[13]This is admittedly a delicate issue since Bolzano's conception of analyticity and syntheticity evolved throughout his writings. As noted, Bolzano in *BD* defines "analytic" in terms of concept-containment, and defines "synthetic" as being not analytic. However, in *WL*, Bolzano defines "analytic" according to a variable term criterion. See Chapter 5 of Lapointe (2011) for more on Bolzano's mature conception of analyticity.

[14]See Kant (1781, B182–3).

diagrams. Hence, instead of proceeding through visualization, and hence, intuition, Příhonský claims that one can just infer from these algebraic tools and the "correct definition of space" in order to arrive at geometric truths.

In order to have cognition of claims about magnitudes as such (the subject matter of algebra for Kant), one needs to construct magnitudes in pure intuition. Again, for Kant, a priori construction is meant to account for the immediate certainty of algebraic axioms, and for drawing synthetic inferences. Thus, an appeal to analytic geometry does not necessarily show that intuition is inessential for mathematical knowledge.

In sum, it is difficult to see how an appeal to the associativity of addition on the natural numbers nor analytic geometry can establish that intuition is eliminable from mathematical knowledge. And in general, any purported counterexample that leaves unexplained the "correctness" of certain definitions or (synthetic) forms of inference fails to demonstrate that intuition is dispensable from mathematical knowledge since intuition is the very thing that witnesses such correctness, according to Kant.

4.2 Bolzano's Theory of Propositions

In the previous section, I argued that neither Bolzano nor Příhonský were able to generate convincingly genuine counterexamples to Kant's account of mathematical knowledge. In this section, I look at an alternative strategy of undermining Kant's claim that intuition plays an indispensable role in securing mathematical knowledge. This strategy consists in putting forward an account of mathematical cognition whereby mathematical cognition (of the sort Kant describes) can be secured by purely conceptual means. Such a strategy is pursued by Bolzano extensively in *WL* by way of his theory of propositions.

In short, Bolzano argues as follows. According to Bolzano, the propositions of mathematics are of a certain type, namely they are *conceptual truths*; they are truths the constituents of which consist solely of concepts. Bolzano argues that the ground of a conceptual truth can only be further conceptual truths, and the case is no different in mathematics. Even if intuition can help guide an epistemic agent to mathematical knowledge, its genuine ground is not found in intuition but in a further conceptual truth. Thus, for Bolzano, intuition is entirely dispensable.

To better understand this argument for the dispensability objection, it is important to note that Bolzano not only questioned the coherence of Kant's philosophy of mathematics, but he questioned the fundamental notions on which it was built. Bolzano found Kant's formulation of terms such as "intuition," "concept," "a priori," "a posteriori," "necessity," and "generality" problematic because Kant defined such terms with reference to our cognitive capacities. Indeed, in the Appendix to *BD*, Bolzano argues that the notion of "a priori intuition" is hopelessly obscure because Kant is working with problematic notions of "a priori" and "a posteriori." For Kant, the distinction between a priori and a posteriori cognitions is based on how a given cognition arises (i.e., whether it arises from experience or not). However, under this conception, judgments are not intrinsically a priori or a posteriori. For instance, Kant's definitions allow one to have empirical cognition of the proposition that "all bachelors are unmarried" by observing a group of bachelors in a room and asking each of them whether he is a bachelor or not. However, because the concept <unmarried> is contained in the concept <bachelor>, one can also have an a priori cognition of the same proposition.

For Kant, then, whether a proposition is a priori or not depends on how the proposition is cognized. In Bolzano's eyes, the problem with Kant's definitions is that they do not pick out *intrinsic* properties of propositions but rather *subject-relative* ones. If the highest kind of mathematical knowledge consists in knowing a mathematical proof that reflects objective relations of ground and consequence and in knowing that it is a grounding proof, then Kant's subject-relative definitions simply won't do. Knowledge of mathematics should not focus on the psychological story on how cognitions are formed. Rather, an account of mathematical knowledge should primarily concern itself with objective, that is, mind-independent ground-consequence relations that obtain between propositions. Bolzano and Příhonský's dissatisfaction with Kant's definitions is captured in the following passage:

> It almost seems as if [Kant] had wanted to make the determination of whether a cognition is *a priori or a posteriori* dependent upon the merely contingent circumstance of how we gain the cognition, whether in the course of experience or through mere reflection. But how unsteady is such a distinction! How uncertain when it comes to making a decision on the question of whether a cognition belongs to one or the other of the two classes! Must one not admit that we reach most, and to a certain extent, all of our cognitions, even the *a priori cognitions*, by means of experience? (Příhonský 1850, p. 47)

At the same time, this is not to say that Bolzano was not concerned with the epistemic agent. In fact, it is the epistemic agent's duty to discover the genuine grounds of a proof rather than satisfy herself with intuitions that merely confer subjective confidence in a proposition's being true.

According to Bolzano, mathematical cognition and representational entities are better understood in relation to what he calls *propositions in themselves* (i.e., "Satz an Sich"), which are the objective mind-independent contents or intentions of linguistic assertions. If there is no ambiguity, I will call "propositions in themselves" simply "propositions." According to Bolzano, the proposition expressed by a linguistic assertion differs from its referent, which is something like a state of affairs.

Representations, which Bolzano calls "ideas in themselves," are the components of propositions. And like propositions, representations are mind-independent entities that differ from referents. Bolzano defines intuitions as simple and singular representations. For a representation to be simple means that it cannot be broken down into further component representations. To be singular means that the representation has only one referent. Concepts are defined to be representations that are not intuitions and have no intuitions as parts.

Bolzano's distinction between a priori and a posteriori falls out of the above definitions. According to Bolzano, an *a priori proposition* is a proposition that has only concepts as components. In contrast, an *a posteriori proposition* is a proposition that contains an intuition. Naturally, Bolzano refers to a priori propositions as “conceptual propositions”, and a posteriori propositions as “empirical propositions”. A *conceptual truth*, then, is just a true conceptual proposition.

Note that these definitions differ from Kant’s. For Kant, representations are mind-dependent entities of which we are immediately acquainted and with which we use to form judgments. Intuitions are particular representations that refer immediately to an object, while concepts are general representations that refer to an object via their marks.

As one might expect, Bolzano found his definitions to be an improvement on Kant’s:

> Our distinctions have the advantage of being thoroughly objective ones, ones which do not rest upon certain external circumstances but rather are grounded in an internal attribute that depends on the object itself and which already pertains to the proposition in themselves, not merely to their appearances in the mind (to judgements and cognitions). (Příhonský 1850, pp. 32–33)

Thus, Bolzano’s definitions purport to provide a better alternative to Kant’s. Instead of focusing on mind-dependent phenomena, Bolzano thought that a more promising epistemology of mathematics could be given if one relied on mind-independent semantic entities and the relations that hold among them.

Bolzano’s theory of propositions also shows why Kant’s philosophy of mathematics gets off on the wrong foot. As mentioned, mathematical truths for Bolzano are *conceptual truths*. They concern truths that consist solely of concepts in Bolzano’s sense. As such, they can only be grounded in further conceptual truths. Thus, as many commentators have concluded, intuition cannot ground mathematical knowledge and can be dispensed with by knowing a proposition’s genuine ground.[15]

I argue that Bolzano’s reasoning is problematic. The inference from the claim that mathematics consists of conceptual truths to the claim that one can and ought to dispense with intuition is too quick: it equivocates subjective with objective grounds—a distinction that Bolzano takes to be of central importance. By conflating subjective grounds with objective grounds, one misunderstands the role intuition plays in mathematical inference. For Kant, intuition was never meant to provide justification for a judgment or ground in Bolzano’s sense. Intuition is not intended to play the role of X in “I perceive X, therefore I know Y” where Y is some mathematical proposition. Intuition is meant to underwrite immediate knowledge of axioms and inferences that go beyond traditional logic.

To think that Bolzano could appeal to the relation of ground-consequence as a way to show the untenability of Kant’s doctrine of a priori construction is misguided and most likely based on a misunderstanding of the role that a priori construction plays in mathematical cognition and inference. In fact, one can see how Bolzano’s ideal of objective proofs in mathematics is *compatible* with Kant’s account. As Bolzano is well aware of, the ground of cognition can be radically different from the ground of truth.

[15] See, for example, Lapointe (2011, p. 17) and Rusnock (1999, pp. 405–6).

Furthermore, an appeal to ground and consequence leaves unexplained how Bolzano would account for our knowledge of ground-consequence relations and ungroundable grounding propositions, i.e., axioms. As Kant has argued, knowledge of either outstrips the logical power of general logic and concept-containment relations.

According to Bolzano, axioms are composed of simple concepts, which he calls "axiomatic concepts". One might think that Bolzano could explain our knowledge of axioms by appealing to our grasp of axiomatic concepts. However, even then, Bolzano would have to provide an explanation of which combinations of axiomatic concepts constitute a genuine ground and which do not. If axiomatic concepts are truly simple, then it is not clear how they alone could disclose such information.

Resisting pure intuition at all costs, Bolzano offers an alternative account of the missing variable that underlies not only synthetic a priori judgments such as axioms but inferred cognition as well:

> Nothing, I say, but that the understanding *has* and *knows* the two concepts *A* and *B*. I think that we must be in a position to judge about certain concepts merely because we have them. For, to say that somebody has certain concepts surely means that he knows them and can distinguish them. But to say that he knows them and can distinguish them means that he can claim something about one of them which he would not want to claim about the others [...] (Bolzano 1837, §305)

Without further details from Bolzano, it is not clear exactly how the "having" or "knowing" of concepts puts one in a cognitive position to go beyond their mere contents and form synthetic a priori judgments, and in particular, to know that a candidate for an axiom is a genuine axiom.

Even if Bolzano's theory of propositions is unable to provide a purely conceptual explanation of how we come to know axioms as grounds, there might still be good reason to think that Kant overestimated the extent to which intuition was needed to secure *mathematical inference*.

In both *WL* and *BD*, Bolzano claims that Kant goes wrong in thinking that without intuition, the only discoverable logical relationships between judgments are the ones given by Aristotelian syllogistic relationships and concept-containment relations. Indeed, as Bolzano emphatically remarks:

> It seems to me therefore that one of Kant's literary sins was that he attempted to deprive us of a wholesome faith in the perfectibility of logic through an assertion very welcome to human indolence, namely, that logic is a science which has been complete and closed since the time of Aristotle. It seems to me that it would be much better to assert as a kind of practical postulate that faith in the perpetual perfectibility not only of logic but of all science should be maintained. (Bolzano 1837, §9)

Bolzano ambitiously sought to go beyond the limits on cognition set forth by the first *Critique*, attempting to redefine logic in such a way that allowed one to uncover the objective dependence relations that held among the truths of mathematics without recourse to intuition.

The force of Bolzano's dispensability objection to Kant can then be salvaged by pointing out that, at least, inferences from known truths did not require the use of intuition. In other words, Bolzano's dispensability objection can be understood as an argument that Aristotle's logic was nowhere near exhaustive and that Kant had mistakenly underestimated our capacity to draw reliable inferences without intuition.

On this point, it is difficult not to side with Bolzano as the development of logic has arguably shown Aristotelian logic to be, in fact, not exhaustive. However, to assess whether Bolzano's dispensability argument is truly a knock-down argument against Kant, more needs to be said concerning why Bolzano took Aristotelian logic to be incomplete. To this, Bolzano simply says:

> There are forms of inference which tell us how to derive certain propositions from others which cannot be derived from them by any syllogisms, no matter how often they are repeated. (Bolzano 1837, §262)

Bolzano has several examples in mind. For instance, Bolzano claims that the inference from "All *A*s are *B*s" and "All *B*s are *A*s" to "Whatever is *A* or is *B*, is *A* and *B*" cannot be reduced to a syllogism. One might be tempted to think that such inferences, if truly irreducible to syllogisms, would already refute Kant's claim that Aristotle's logic was complete. Hence, such inferences would open the possibility of synthetic inferences that did not require the use of intuition.

In *BD* and especially *WL*, Bolzano attempts to systematize various syllogistic and non-syllogistic forms of inference. While in *BD*, Bolzano admits that "on the other hand, how propositions with simple concepts could be proved other than through a syllogism, I really do not know" (Bolzano 1810, §20), he concedes that the inference from "this is a triangle" to "this is a figure the sum of whose angles equals two right angles" requires knowledge of extra-logical facts (i.e., the general fact that "the sum of the angles in any triangle equals two right angles"). Indeed, Bolzano remarks that a number of his inference rules are such that their "validity or invalidity ... can be assessed only if we have knowledge of matters outside of logic" (Bolzano 1837, §223).

Recall that for Kant, general logic concerns the "logical form in the relation of cognitions to one another" (Kant 1781, B76–9). As I have pointed out, Kant's characterization of general logic does not necessarily entail, as Kant thought, that the principles of logic are exhausted by Aristotle's syllogisms. Regardless of what one takes as the principles of logic, what matters for Kant is that they concern the mere form of thinking. Thus, I take it that Kant's critical project is not essentially tied to Aristotelian logic.[16]

Insofar as Bolzano's proposed non-syllogistic inferences are sufficiently formal, Kant would not disagree with Bolzano. However, as mentioned before, many of Bolzano's proposed inferences admittedly do not possess this feature and require the use of non-logical facts. To use the above example, the inference from "this is a triangle" to "this is a figure the sum of whose angles equals two right angles," requires knowing the general fact that "the sum of the angles in any triangle equals two right angles." A Kantian presumably has no problem in permitting forms of

[16] This is not to say that serious complications will not arise if Kant's critical philosophy is adapted to a different set of logical laws.

inference that presuppose knowledge of non-logical facts as counting as genuinely part of logic. In fact, he calls these inferences as being part of applied logic, which is distinct from general logic because the former concerns inferences of a particular subject matter. However, the point is that a truly logical inference involves the mere form of thinking without resorting to non-logical facts.

Neither does an appeal to Bolzano's celebrated notion of deducibility (i.e., "Ableitbarkeit") get at a viable alternative to Kant's intuition in the case of securing mathematical inference (Bolzano 1837, §154). According to Bolzano, one can arrive at (some) reliable forms of inference by taking an argument, letting certain terms of the argument be variable, and testing to see if every substitution instance that makes the premises true makes the conclusion true. One might contend that such a method isolates certain forms of reliable inference, which can be used to eliminate intuition from the inference to a proposition from its ground. However, the problem with this strategy is that Bolzano still does not have the resources to explain cognition of grounds since not every instance of a deducibility relation obtaining between sentences, say, *P* and *Q* implies that there is a grounding relation obtaining between *P* and *Q*. In modern parlance, not every case of logical validity implies metaphysical grounding or explanation.[17] Furthermore, Kant could argue that choice of constant and variable terms (unless one takes all non-logical terms as variable) is a choice that violates the formality of logic. Even if Bolzano rejects Kant's conception of logic as formal, Bolzano has to concede that in order to mechanically carry out a deducibility test, one has to have prior cognitive access to a potentially infinite class of truths, the cognition of which we either do not have as finite epistemic subjects or one of which remains in need of explanation.

Ultimately, Kant has at his disposal intuition to explain cognition of both ground-consequence relations and grounds. Such knowledge is arguably outside of the scope of Bolzano's objective definitions, and his notion of ground-consequence and deducibility. These resources alone are unable to provide a substitute for Kant's intuition, and thus, the dispensability objection loses its argumentative force.[18]

In sum, I hope to have shown that the particularity of intuitions does not preclude them from enjoying a necessary role in mathematical cognition. Furthermore, neither does an appeal to Bolzano's more "objective" definitions get at problematic features of Kant's account nor do they provide resources that can act as an alternative

[17] One can give the following counterexample. Given any proposition *P*, *P* logically implies itself. However, not every proposition *grounds* itself.

[18] At the end of the day, I argue that the tension between Kant's and Bolzano's respective definitions of terms such as "intuition" and "a priori" is merely apparent. I see no reason why a Kantian has to reject Bolzano's definitions as long as he could supplement them with her own. For Kant, a mathematical inference is composed of a grounding cognition and an a priori construction that allows for the possibility of genuine cognition from the ground to the inferred proposition in a way that goes beyond the laws of traditional logic. Kant's account leaves room for what Bolzano would consider to be an objective ground of a mathematical truth as its objective ground. As I see it, Kant's account does not necessarily impose prohibitions on mathematical practice in a way that would preclude any instances of a ground-consequence from being genuine cases of ground-consequence.

to Kant's account. Even if our knowledge of mathematics ought to follow from concepts without an appeal to intuition, Bolzano has yet to provide a compelling explanation as to how that is possible.

While various commentators have attributed Bolzano with undermining Kant's doctrine of pure intuition on the basis of mathematical definitions and new logical notions, I have argued that if Bolzano undermines Kant's doctrine of intuition, it is not Bolzano's alternative definitions nor his logical notions of grounding and deducibility that do the job.

5 Concluding Remarks: A Deeper Disagreement

To conclude, I would like to make some remarks concerning why Bolzano felt generally dissatisfied with Kant's critical philosophy. I argue that many of Bolzano's objections stem from his rejection of the claim that a "critique of pure reason" is necessary for securing and explaining reliable forms of inference and claims to a priori knowledge. While Kant thought he was performing a service for philosophy, which simultaneously preserved and explained the a priori status of mathematics, Bolzano saw the conclusions of the critique as inhibiting mathematical progress. The following passage summarizes this sentiment:

> The path therefore which the *Critique* pursues for deciding the question about the validity of our cognition is not at all the right one. The faculty of cognition of human beings or of thinking beings in general must not be analyzed first. Rather it is necessary to look into the nature of the truths in themselves [. . .] Only after this has happened can one usefully deal with cognizing and the conditions of cognizing. (Příhonský, 1850, p. 146)

Bolzano saw a critique of reason to be unnecessary because we can be sure of the reliability of our inferences and our claims to a priori cognition by studying general features of propositions in themselves. The nature of Bolzano's project reveals that he was not discouraged by Kant's conclusion that our theoretical synthetic a priori cognition cannot outstrip what is given to us.

While Kant or a Kantian would not necessarily disagree with Bolzano's view that there is an objective realm of propositions standing in ground-consequence relations, one of the central conclusions of the first *Critique* is that there are certain limitations on our ability to cognize such relations. Namely, our cognition is limited to objects of possible experience. From a Kantian point of view, because Bolzano seeks to extend the use of reason beyond possible experience, Bolzano's project runs the risk of problematically extending our claims of a priori cognition beyond reason's proper bounds, making reason vulnerable to the endless metaphysical disputes that motivated Kant's *Critique* in the first place. The naive set theory (which features an arguably inconsistent unrestricted comprehension axiom[19]) developed in

[19] See Bolzano (1851, §14).

Bolzano's 1851 *Paradoxien des Unendlichen*[20] might well be a perfect example of the unfortunate consequences that may result from reason extending its claims of cognition beyond its proper bounds.

If human agents with particular cognitive capabilities are the ones carrying out Bolzano's project as he envisions it, then it seems that any account regarding the secure advancement of knowledge needs to make some essential reference to those cognitive capabilities. Before we can engage in the type of foundational reform that Bolzano suggests, it seems necessary to gauge the extent to which we are epistemically and mechanically capable of tracking objective dependence relations at all.

Acknowledgements Thanks are owed to Karl Ameriks, Mic Detlefsen, Matteo Bianchetti, Curtis Franks, Eric Watkins, two anonymous referees, and the organizers and participants at CSHPM 2016 for their helpful comments and discussion.

References

Aristotle (1993) Posterior Analytics, 2nd edn. Clarendon Press, Oxford

Bolzano B (1810) Contributions to a Better-Grounded Presentation of Mathematics. In: Russ S (ed) The Mathematical Works of Bernard Bolzano, Oxford University Press, pp 83–139

Bolzano B (1837) Theory of Science. Oxford University Press, Oxford

Bolzano B (1851) Paradoxes of the Infinite. In: Russ S (ed) The Mathematical Works of Bernard Bolzano, Oxford University Press, pp 591–670

Coffa A (1991) The Semantic Tradition from Kant to Carnap: To the Vienna Station. Cambridge University Press, Cambridge

Friedman M (2010) Kant on geometry and spatial intuition. Synthese 186(1):231–255

Heis J (2014) Kant (vs. Leibniz, Wolff and Lambert) on real definitions in geometry. Canadian Journal of Philosophy 44(5–6):605–630, doi:10.1080/00455091.2014.971689

Kant I (1781) Critique of Pure Reason. Cambridge University Press, Cambridge

Lambert JH (1761) Treatise on the Criterion of Truth. In: Watkins E (ed) Kant's Critique of Pure Reason: Background Source Material, Cambridge University Press, Cambridge, pp 233–257

Lapointe S (2011) Bolzano's Theoretical Philosophy An Introduction. Palgrave Macmillan, Basingstoke

Příhonský F (1850) New Anti-Kant. Palgrave Macmillan, Basingstoke

Rusnock P (1999) Philosophy of mathematics: Bolzano's responses to Kant and Lagrange. Revue d'histoire des sciences 52(3/4):pp. 399–427

Russell B (1919) Introduction to Mathematical Philosophy. George Allen & Unwin, Ltd., London

[20] "Paradoxes of the Infinite."

Poincaré on the Foundation of Geometry in the Understanding

Jeremy Shipley

Abstract This paper is about Poincaré's view of the foundations of geometry. According to the established view, which has been inherited from the logical positivists, Poincaré, like Hilbert, held that axioms in geometry are schemata that provide implicit definitions of geometric terms, a view he expresses by stating that the axioms of geometry are "definitions in disguise." I argue that this view does not accord well with Poincaré's core commitment in the philosophy of geometry: the view that geometry is the study of groups of operations. In place of the established view I offer a revised view, according to which Poincaré held that axioms in geometry are in fact assertions about invariants of groups. Groups, as forms of the understanding, are prior in conception to the objects of geometry and afford the proper definition of those objects, according to Poincaré. Poincaré's view therefore contrasts sharply with Kant's foundation of geometry in a unique form of sensibility. According to my interpretation, axioms are not definitions in disguise because they themselves implicitly define their terms, but rather because they disguise the definitions which imply them.

Keywords Henri Poincaré • Non-Euclidean geometry • Geometry • Axiomatics • Definition

---

This work was supported by funding from the Lorraine Region and much of the research was conducted at the Archives Henri Poincaré at the Université de Lorraine. Early versions of the paper were presented to seminars at the Université de Lorraine and Universit at Konstanz in 2013. Subsequent versions were presented at the Midwest Philosophy of Mathematics Workshop in 2015, the Canadian Society for the History and Philosophy of Mathematics Annual Meeting, the Lehigh University Annual Conference in Philosophy, and the Biennial Meeting of the Philosophy of Science Association in 2016. I would like to thank Gerhard Heinzmann, Philippe Nabbonaud, Jean-Pierre Ferrier, Bruno Blind, Denis LeLarge, Jacintho Del Vecchio, Dina Emundts, Jacob Rosenthal, Jochen Briesen, Amber Griffioen, Gregory Landini, Colin McLarty, and Jamie Tappenden for helpful discussions of the topic of this paper.

J. Shipley (✉)
Volunteer State Community College, Gallatin, TN 37066, USA
e-mail: jeremy.shipley@volstate.edu

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*, Proceedings of the Canadian Society for History and Philosophy of Mathematics La Société Canadienne d'Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_2

1 Introduction

In this paper I will offer an interpretation of Henri Poincaré's views on the foundations of mathematics, revolving around a revised assessment of his understanding of geometry. My interpretation of Poincaré on the foundations of geometry departs from the established view, according to which he held that axioms implicitly define geometric terms. In the following section I will present the established view and briefly summarize the textual evidence for it. Subsequently I will explain and provide evidence for my revised view. According to the established view, Poincaré regarded the axioms of geometry as uninterpreted schemata, providing implicit definitions of geometrical terms. I will contend that Poincaré in fact maintained that the axioms of geometry are properly understood as assertions about invariants of geometric groups.

This paper extends and clarifies the interpretation of Poincaré's philosophy of geometry that has recently been articulated by David Stump and Gerhard Heinzmann. In particular, I disagree with part of Stump and Heinzmann's *Stanford Encyclopedia of Philosophy* article on Poincaré; viz., the claim that "concerning pure and applied geometry, Poincaré holds the modernist-sounding view that we have no pre-axiomatic understanding of geometric primitives" (Heinzmann and Stump 2014, §3.2).[2] I will argue that according to Poincaré we do indeed have a pre-axiomatic understanding of geometric primitives. The revised view that I will defend aligns well with Michael Friedman's interpretation of Poincaré's conventionalism, which emphasizes the group theoretical perspective that is key to comprehending Poincaré's comments on geometry (Friedman 1997). The paper contributes to the literature primarily by directly addressing the established view, articulating an alternative, and showing how that alternative is better supported; secondarily, I will clarify the relationship of Poincaré's views to those of his contemporaries by showing that, contrary to Russell's perception, the Poincaré-Russell polemics concerned not whether geometrical primitives should receive nominal definitions but rather what are the primitives, and also by emphasizing the differences between Poincaré and Hilbert.

2 The Established View

According to the established view, Poincaré understood geometric axioms to be uninterpreted schemata that define mathematical concepts without reference to intuited objects. Accordingly, many have thought that Poincaré held that geometrical concepts are to be defined by implicit definition. This view of Poincaré can be found

---

[2]My interpretation, however, is strongly influenced by Heinzmann, and in particular on his view of Poincaré's account of mathematical reasoning as founded on the "capability to follow an action schema" (Heinzmann 1998, p. 47).

expressed by the logical positivists, and it influenced their understanding of his conventionalism. Moritz Schlick wrote:

> To define a concept implicitly is to determine it by means of its relations to other concepts. But to apply such a concept to reality is to choose, out of the infinite wealth of relations in the world, a certain group or complex and to embrace this complex as a unit by designating it with a name. By suitable choice it is always possible under certain circumstances to obtain an unambiguous designation of the real by means of the concept. Conceptual definitions and coordinations that come into being in this fashion we call conventions (using this term in the narrower sense, because in the broader sense, of course, all definitions are agreements). It was Henri Poincaré who introduced the term convention in this narrower sense into natural philosophy; and one of the most important tasks of that discipline is to investigate the nature and meaning of the various conventions found in natural science (Schlick and Blumberg 1974, p. 71).

Rudolf Carnap, too, mentioned Poincaré's philosophy of science in the context of his discussion of "structural definite descriptions," which he relates to Hilbert's implicit definitions, suggesting that he too viewed Poincaré's conception of geometry as aligned with the idea of implicit definitions (Carnap 1969, p. 27–30).[3] According to the established view, convention enters into science through the coordination of axiom schema to reality.

More recently, Shaprio, in his book defending *in re* structuralism, situates Poincaré as a proponent of implicit definitions of geometrical concepts by uninterpreted axiom schemata (Shapiro 2000, p. 154–156). Shapiro draws his interpretation from the influential work of Alberto Coffa, who parallels the Poincaré-Russell polemics on the foundations of geometry with the Hilbert-Frege correspondence, by aligning Poincaré and Hilbert on the satisfactoriness of implicit definitions and Russell and Frege on the demand for nominal definitions (Coffa 1993). While it is true that Russell understood Poincaré to be rejecting the need for nominal definitions of geometric primitives, on my view this is a misunderstanding on Russell's part: Poincaré was arguing that Russell had the wrong primitives and not that there wasn't any need for primitives. Here is how Coffa presented the basic issue at stake in that debate:

> [Russell's] reasoning [against Poincare] involved an appeal to a principle that may be called the thesis of semantic atomism. . . This principle says that if a sentence S is to convey information (or, as Russell or Frege would put it, to express a proposition), then its grammatical units must have a meaning before they join their partners in S (Coffa 1993, p. 131)…
> Poincare's conventionalism was based on the idea that in order to understand geometry, one must stand Russell's argument on its head: Since geometric primitives do not acquire their meaning prior to their incorporation into the axiomatic claims, such axioms do not express propositions in Frege's or Russell's sense (Coffa 1993, p. 133).

[3]However, Carnap views Poincaré as not having gone far enough, since, according to Carnap Poincaré conceived of science as concerned with relations rather than taking the further step of saying science is concerned with the Russellian structure-type of a relation.

To put a fine point on it, I will show that according to Poincaré the geometric "primitives" in question do in fact acquire meaning prior to their incorporation in typical axiomatizations, precisely because they are not, after all, primitive.

Though I will be offering an alternative, there can be no doubt that there is a textual basis for following the positivists, Coffa, Shapiro, and others in accepting the established view. The established view is supported by three pillars:

1. The established view provides an appealing and direct explanation of Poincaré's statement that axioms are definitions in disguise.
2. The established view also explains why Poincaré concurs with Hilbert's claim that mathematics is constrained by consistency alone.
3. Following Schlick, the established view coheres with a common understanding of Poincaré's conventionalism.

Defenders of the established view may lean upon Poincaré's repeated insistence that axioms are definitions in disguise, constrained only by non-contradiction, and that they are conventions. How else can this be understood than in accordance with the established view? Accordingly, axioms are definitions in disguise in the sense that they implicitly define the concepts of point, line, plane, etc. by characterizing a relational structure among terms, which then must be coordinated with the world by appropriate conventions.

In *Science and Value* Poincaré argues that applied geometry concerns neither synthetic *a priori* judgments nor experimental facts. It does not concern synthetic *a priori* judgments because it is possible to imagine experiences that would lead us to deny some of the axioms of Euclidean geometry, and it does not concern experimental facts because it is an exact science. He concludes that the claims of applied geometry are conventional:

> They are *conventions*. Our choice among all possible conventions is *guided* by experimental facts; but it remains *free* and is limited only by the necessity of avoiding all contradiction. Thus it is that the postulates can remain *rigorously* true even though the experimental laws which have determined their adoption are only approximate.
> In other words, *the axioms of geometry* (I do not speak of those of arithmetic) *are merely disguised definitions* (Poincaré 1905, p. 65)

Understood in the following way, this characterization of the axioms of geometry accords with a view commonly associated with Hilbert. Axioms are possible conventions because they are schematic and are to be coordinated to empirical reality by conventional definitions. A "line" may be a rail track or a ray of light, and a "point" may be a rail station or a star.[4] Using the inferential relations encoded in the axioms, we form abstract mathematical concepts of point, line, etc. as places in an abstract structure, but there is no privileged, intuitive interpretation according

[4]Hilbert is credited with having said of geometric terms that "One must be able to say 'tables, chairs, beer mugs' each time in place of 'points, lines, places'," (Blumenthal 1935, p. 402–403).

to which the truth or falsity of the axioms is to be assessed. Accordingly, axioms provide implicit definitions of general concepts, and are only apparent assertions about objects.[5]

In the context of the established view, good sense is made of the fact that Poincaré, like Hilbert in his correspondence with Frege, emphasized that consistency was the only constraint on mathematics, stating that "a mathematical entity exists, provided its definition implies no contradiction, either in itself, or with the propositions already accepted" (Poincaré 1905, p. 60). Such passages may incline one to assimilate Poincaré's views on the foundations of geometry to Hilbert's and to analogize the Hilbert-Frege and Poincaré-Russell polemics. For, the claim that mathematics is constrained only by consistency may seem naturally to be understood in a framework that prioritizes axiomatization over genetic intuitions, because a property naturally ascribed to sets of sentences or propositions, viz. consistency, is given priority over intuition as a criteria of existence.

Lastly, it may also be contended that Poincaré's famous non-falsifiability arguments for geometrical conventionalism are only to be understood in the context of the established view. Any axiom system for geometry can be adopted and held to come what may, making compensating changes to other theories should anomalous observations occur. Accordingly, geometrical axioms must be conventions because of their non-falsifiability in practice. The relational structure encoded in the implicit definition is non-falsifiable for the system to which the definitions are coordinated, and hence it is conventional.

3 A Revised Interpretation

Notwithstanding the support available for the established view, I will argue in what follows that it leads us to neglect important distinguishing features of Poincaré's view of the foundations of geometry: that the sense in which axioms are disguised

---

[5]A great deal has been written about Hilbert's view in his *Grundlagen der Geometrie*, the relationship between this view and the modern model theoretical understanding of axioms, and Frege's response to Hilbert in their correspondence. A reviewer points out that an appeal to Fregean second-order concepts can accommodate Hilbert's method, because axioms, understood as implicit definitions, define classes of structures. This proposal is discussed in William Demopolous's essay "Frege, Hilbert, and the Conceptual Structure of Model Theory" (Demopoulos 1994, p. 219). Patricia Blanchette has argued that Frege, while failing to adequately appreciate Hilbert's method and the new sense of 'consistency' that method involved, does succeed in articulating a sense in which Hilbert fails to demonstrate the consistency of denying the parallel postulate: viz., when that postulate is understood as expressing a Fregean thought (Blanchette 1996). The motivation for Frege's resistance to Hilbert's approach, and to regarding geometry as a branch of logic like arithmetic, has been traced to his experience with 19th century geometry, beginning with Wilson (1992) and further developed in Tappenden (1995). I discuss Frege's position that geometry is semantically founded on reference to particulars presented in sensible intuition in my "Frege on the Foundation of Geometry in Intuition" (Shipley 2015).

definitions differs for Poincaré from Hilbert's sense, that Poincaré's conventionalism can only be restricted to geometry (rather than implying a more radical Duhem-Quine style holism) if the established view is rejected, and that Poincaré's and Hilbert's equal emphasis on consistency is circumstantial.

The central feature of the established view is the treatment of geometrical axioms as schematic rather than assertoric. According to the revised interpretation of Poincaré's views on geometry that I wish to advance, geometric axioms are assertions about points and lines, which are defined prior to axiomatization by what Poincaré calls "invariant" sub-groups (in modern parlance: "normal" sub-groups). This interpretation allows us to see Poincaré's views on the foundations of geometry as intimately related to his mathematical practice and to his emphasis on the importance of group theory. While the established view may be fully articulated without mention of group theory, the revised interpretation gives central place to Poincaré's insistence that geometry is the study of the properties of certain groups.

Poincaré's claim that geometry is the study of a group makes its first appearance in his reputation-making mathematical work on Fuchsian functions, produced in 1880. In a supplement to the prize essay submitted to the *Académie des Sciences* in Paris he states:

> It [geometry] is the study of the group of operations formed by the displacements to which one can subject a body without deforming it. In Euclidean geometry the group reduces to the rotations and translations. In the pseudogeometry of Lobachevski it is more complicated (Poincaré 1997, p. 14).

The primary concern of this work is differential equations rather than geometry and its foundations, and these early comments do not receive full philosophical articulation. Geometry arises in the course of studying and classifying functions that are solutions to second-order differential equations having specified properties.[6] The key to Poincaré's discoveries in the field of Fuchsian functions was his realization that the transformations employed in the definition of Fuchsian functions form a group that is isomorphic to the group employed in the definition of the hyperbolic geometry of Lobachevsky, leading him to the analogy "the Fuchsian function is to the geometry of Lobachevsky what the doubly periodic function is to that of Euclid" (Poincaré 1997, 15). In a text originally composed in 1901 and later published in *Acta Analytica* in 1921, Poincaré emphasizes the fundamental role of group theory by asserting an analogy between his work on solutions to classes of differential equations and Galois' work on the solution of polynomials by radicals (Poincaré 1983, 259). According to the interpretation that I am arguing for, Poincaré's philosophical commitments in the foundations of geometry are intimately related to his formative mathematical experience and to a broader view of mathematics.

Groups play a fundamental role in Poincaré's more philosophical account of the conventions we employ in order to categorize and control the changes that occur in our perceptual stimulus, presented in his 1898 article "On the Foundations of

[6]For details see John Stillwell's helpful introduction and the primary texts (Poincaré 1985) as well as Jeremy Gray's commentary accompanying the further texts (Poincaré 1997).

Geometry" published in *The Monist*. The form of a group of operations that is imposed in categorizing sensible changes as displacements does not, in Poincaré's view, depend in any way on the qualitative character of the sensible changes. It is, rather, strictly formal and its properties can be studied formally. Poincaré writes:

> We have now to study the properties of the group. These properties are purely formal. They are independent of any quality whatever, and in particular of the qualitative character of the phenomena which constitute the change to which we have given the name displacement. We remarked above that we could regard two changes as representing the same displacement, although the phenomena were quite different in qualitative nature. The properties of this displacement remain the same in the two cases; or rather the only ones which concern us, the only ones which are susceptible of being studied, are those in which quality is no wise concerned (Poincaré 1898, p. 13).

Continuing, Poincaré defines and discusses the notion of isomorphism, and the idea that the formal properties of a group are what are studied by mathematics. For example, the group of permutations of five cards is isomorphic to the group of permutations of five marbles; the qualitative difference between cards and marbles is irrelevant to mathematics. Similarly, all that matters to mathematical geometry are the formal properties of the group of displacements.

The structure of a group affords the definitions of the straight line and of the point given in the *Monist* article. Poincaré arrives at the notions of "point" and "line" in an entirely group theoretically driven way. The rotative sub-group is prior in conception to the point, which is conceived only as that which is fixed by the rotation, and the rotative sheaf, defined by taking the common displacements of two rotative groups, is prior in conception to the line, which is conceived as the axis that is fixed by a rotation: "Here [with the rotative sheaf] is the origin of the notion of the straight line, as the rotative sub-group was the origin of the notion of the point." (Poincaré 1898, p. 20). Poincaré insists that although the idea of a rotative group is suggested by sensation, it is conceived in a purely formal manner by the understanding, so that our reasoning may become precise.[7] There is a genuine ambiguity in the text. Poincaré refers to the rotative sub-group and sheaf as, respectively, the "origin" of our notions of point and line. I will not take a position on how to interpret this. On one interpretation the points and lines just are the sub-groups. According to another interpretation, points and lines are parts of an amorphous, continuous manifold, which are invariant under the action of the sub-groups in question.[8]

[7] This repeats a theme in Poincaré's work, wherein vague, inaccurate, or even inconsistent ideas are suggested by sensation then made precise and consistent through mathematical reflection. His discussion of Fechner's account of the subjective continuum and its precisification in mathematical understanding, for example, follows this pattern (Poincaré 1905, p. 22).

[8] Because of this ambiguity, and because I take no position on the correct interpretation, I have not commented on the status of Poincaré's *analysis situ* in this paper. However, there is evidence in his review of Russell's *Essay on the Foundations of Geometry* that in the 1890s he conceived of *analysis situ*, like metrical and projective geometry, as founded on a group of transformations (Poincaré 1899, §19).

4 Definitions in Disguise

This brings us to the first pillar of support for the established view. What sense can be made of Poincaré's claim that geometrical axioms are "definitions in disguise" if the established view is abandoned? Is there a tension between the view of axioms as definitions in disguise and the definition of points and lines as group invariants? In fact, keeping in mind that Poincaré gives prior and independent definitions of the primitive terms occurring in geometric axioms can help us to better understand what he means when he says that axioms are definitions in disguise, and how his intended meaning differs from the view that has been attributed to him.

For Poincaré, the claim that axioms are definitions in disguise does not mean that the axioms are themselves definitions that are disguised as assertions. Rather, axioms are definitions in disguise because they disguise the definitions that imply them. One reason to adopt this interpretation is that it allows us to make the claims about axioms and definitions cohere with the plainly stated definitions of "point" and "line" given in "On the Foundations of Geometry" (the 1898 *Monist* article) and to make sense of the consistently insisted upon priority of group theory. However, the reasons for adopting this interpretation are not exclusively indirect, and in fact there is strong textual evidence for this way of viewing things. When read in full context it is clear that while Poincare does, of course, say that axioms are definitions in disguise, he does not mean that axioms implicitly define mathematical concepts; rather, he means that there are correct definitions that imply the axioms, which are disguised when we treat the axioms as primitive.

In *Science and Hypothesis* the claim that axioms are "disguised definitions" comes in the summation of Ch III, and is presented as an alternative way to state that the axioms of geometry "are neither synthetic *a priori* judgments nor experimental facts," but rather are "conventions" (Poincaré 1905, p. 65). We shall return to the relevance of the revised interpretation that I am arguing for to the proper understanding of Poincaré's conventionalism shortly, but for now I wish to focus attention on passages which precede these widely cited quotations, which I claim will help us better understand the statement that axioms are definitions in disguise.

In a section of Ch III titled "The Implicit Axioms" Poincaré begins with the question "Are the axioms explicitly enunciated in our treatises the sole foundations of geometry?" The answer implied by the ensuing discussion is plainly "no." There are unenunciated "implicit axioms" that are implied by the correct definitions of geometrical entities. It is important to stress that Poincaré holds clearly that there are correct definitions of geometrical entities in this section of *Science and Hypothesis* and that these definitions, which entail the implicit axioms, are the same as those given in "On the Foundations of Geometry":

> "It may happen that the motion of a rigid figure is such that all the points of a line belonging to this figure remain motionless while all the points situated outside of this line move. Such a line will be called a straight line." We have designedly, in this enunciation, separated the definition from the axiom it implies.

> Many demonstrations, such as those of the cases of the equality of triangles, of the possibility of dropping a perpendicular from a point to a straight, presume propositions which are not enunciated, for they may require the admission that it is possible to transport a figure in a certain way in space (Poincaré 1905, p. 62).

Contrary to the view that Poincaré held that the usual axioms provide implicit definitions of geometric concepts, Poincaré is here articulating (less than perfectly clearly, it must be admitted) the same view that he holds in "On the Foundations of Geometry," that we obtain concepts like "point" and "line" from invariant subgroups; these are what determine the possibility of transporting figures. He writes that affirming the existence of a certain group of operations "constitutes the axioms of Euclid" (Poincaré 1898, p. 18). Indeed, I do not think that these passages of *Science and Hypothesis* can support the claim that Poincaré rejects the traditional demand for nominal definitions. Poincaré is, rather, insisting that the nominal definitions that have traditionally been given are inadequate, and need to be replaced by definitions that constitute geometrical objects based on the conventional assumption of a chosen group; this "true definition" is not an implicit definition in the sense of Schlick.

5 Consistency

So much for the first pillar of support for the established view. We have seen that the revised interpretation of Poincaré's foundation of geometry that axioms are definitions in disguise can equally well, if not better, account for the relevant texts. Accordingly, axioms are not definitions in disguise because they themselves implicitly define their terms, but rather because they disguise the definitions which imply them. We turn now to the second support for the established view: viz., that it makes the best sense of Poincaré's concurrence with Hilbert that mathematics is constrained only by non-contradiction. To better understand Poincaré's view on this, we should situate his thinking on the foundations of geometry in relation to Kant's.

A clear agenda of Poincaré's *Monist* article is to vitiate Kant's account of the origin of geometry in a single synthetic *a priori* form of sensibility and to replace that account with a theory of the origin of geometry in the understanding. Importantly, Poincaré's proposal for doing this conceives of forms of the understanding as groups of operations rather than discursive concepts. That Poincaré opposes the account of geometric notions as originating in sensibility, either empirically or transcendentally, is clear from page 1 of the essay:

> Our sensations cannot give us the notion of space. That notion is built up by the mind from elements which pre-exist in it, and external experience is simply the occasion for its exercising this power, or at most a means of determining the best mode of exercising it. Sensations by themselves have no spatial character (Poincaré 1898, p. 1).

The remainder of the essay is devoted to articulating an alternative, based on the claim that "we have within us, in a potential form, a certain number of models of groups, and experience merely assists us in discovering which of these models departs least from reality" (Poincaré 1898, p. 13), which form "the common patrimony of all minds" (Poincaré 1898, p. 18).

Now, on the established view we can well understand how geometry could be founded in the understanding rather than sensibility. Accordingly, geometry concerns relations between what Kant called "discursive concepts," which are constrained only by the law of non-contradiction but which, according to Kant, as empty linguistic forms thus lack "objective validity."[9] Though it is a bit anachronistic, I think it is not entirely misleading to assimilate Kant's "discursive concepts" or "forms of thought" to Schlick's implicit definitions. Hilbert's claim that the only constraint on mathematics is the law of non-contradiction, that in mathematics existence means non-contradiction, can be read in this context as rejecting the Kantian standard of objective validity and that any consistent axiomatization may have a corresponding model freely postulated. Recalling that Poincaré makes similar statements to Hilbert's concerning consistency and existence, ought we not, as the second pillar of the established view has it, interpret him in the same way, and ought we to after all attribute to Poincaré a view that gives axiom schemata and implicit definition a central role?

An alternative, non-discursive, interpretation of Poincaré's view of the role of non-contradiction is available, which better accords with his insistence on group theoretical foundations of geometry and with algebraic methods that Poincaré celebrated as fundamental to the advance of mathematics in the 19th century.[10] I wish to distinguish the question of the consistency of an extension of a system of operations from the question of the consistency of a set of propositions. On the one hand, if we understand the claim that mathematics is constrained only by non-contradiction as applied to consistency in the latter sense, then we are lead naturally to Hilbert's view and to the axiomatic method as a mode of definition of mathematical concepts, which may be considered as discursive concepts in Kant's sense; we just drop the constraint of objective validity and its association with transcendental conditions of sensory representation.

[9]Early in the transcendental aesthetic, Kant denies that space is a "discursive concept" (Smith 1929, B37). According to Kant, the process of geometrical object construction by the productive imagination is governed and restricted by space as a form of sensibility. The formation of concepts in the understanding is restricted only by the law of non-contradiction, but some non-contradictory concepts, such as that of a figure closed by two straight lines, are nevertheless concepts of impossible objects. They lack objective validity. Kant distinguishes between "form of thought," which has a purely logical significance, and the conditions for the possibility and necessity of objects of thought, which concerns experience and its synthetic unity (Smith 1929, B267). To determine that a concept is of a possible object of thought, it is necessary but not sufficient to show that it is non-contradictory. Sufficiency demands, further, that a concept must satisfy the conditions for the possibility of experience. The point is made nicely by Michael Friedman (Friedman 1990, p. 216–218) who references the following passages: (Smith 1929, B271) and (Smith 1929, B298-9). See also (Smith 1929, B151-2) and (Smith 1929, B196) for more on objective validity and the relationship between concepts formed in the understanding and the forms of sensibility through construction by the productive imagination.

[10]See (Poincaré 1983, 259), which is discussed above.

Alternatively, we may view consistency as restraining the freedom we have in postulating extensions of groups of operations. Treating consistency as a constraint on extending systems of operations, rather than on the acceptability of axiom schemata as implicit definitions, allows one to understand the consistency constraint in a way that harmonizes well with Poincaré's description of the development of number systems beginning with basic arithmetic. Therein, we take for granted an understanding of the operation $+1$ and use this understanding to define further operations by recursion, through a "monotonous series of reasonings" (Poincaré 1905, p. 36). It is in this context that Poincaré makes his controversial remarks regarding the non-reducibility of the principle of recursion to the principle of non-contradiction, a claim which he maintained would preclude a complete logicist reduction of mathematics. Our purpose is not to evaluate this claim, however, but to better understand Poincaré's general way of thinking about the foundations of mathematics, and his remarks well accord with the view that he considered consistency to be the constraining factor on the process of extending systems of operations. For it is this constraint that we know to be unviolated by "the power of the mind which knows itself capable of conceiving the indefinite repetition of the same act when once this act is possible" (Poincaré 1905, p. 39).

We may bring this discussion of Poincaré's view of the foundations of arithmetic into connection with his discussion of the operations that form the groups at the foundation of geometry by reflecting on the following passage, which introduces the section of "On the Foundations of Geometry" that is concerned with rotative sub-groups:

> The number of sub-groups [of the group of all displacements] is infinite; but they may be divided into a rather limited number of classes of which I do not wish to give here a complete enumeration. But these sub-groups are not all perceived with the same facility. Some among them have been only recently discovered. Their existence is not an intuitive truth. Unquestionably it can be deduced from the fundamental properties of the group, from properties which are known to everybody, and which are, so to speak, the common patrimony of all minds. Unquestionably it is contained there in germ; yet those who have demonstrated their existence have justly felt that they had made a discovery and have frequently been obliged to write long memoirs to reach their results.
>
> Other sub-groups, on the contrary, are known to us in much more immediate manner. Without much reflexion every one believes he has a direct intuition of them, and the affirmation of their existence constitutes the axioms of Euclid. Why is it that some sub-groups have directly attracted attention, whilst others have eluded all research for a much longer time? We shall explain it by a few examples (Poincaré 1898).

There are two ideas to draw from this passage that speak in favor of the revised interpretation. First, in general support of the revised interpretation, the conceptual priority of groups of operations is affirmed by the claim that the axioms of Euclid are constituted by the existence of a group. Second, in support of the present reflections on the consistency constraint, the passage strongly suggests the foundational importance of research into groups and their sub-group structure, which I would suggest is best understood on the model of Poincaré's view of the foundations of arithmetic in the following way: a "germ" is developed through the repetition and combination of generating operations[11] I elaborate on this idea, and the relationship the way groups may be defined and some of Poincaré's broader commitments regarding proper definition below.

In the light of the role that the consistency constraint has for Poincaré that I have been developing, we can see how consistency can have been important for him without relying on the established view that he took axioms to be schemata and to function as implicit definitions. Specifically, consistency constrains the construction of forms of the understanding by restricting the introduction of generators or relations, either primitively or to extend existing groups.

6 Conventionalism

So much for the second pillar supporting the established view. I turn now to the third pillar: that the established view lends to a nice account of Poincaré's conventionalism. Schlick's interpretation of Poincaré, as we have detailed above, makes natural sense of his conventions in the following way: conventions coordinate terms in implicit definitions to real objects, properties, and relations. The main problem for the established view concerning the interpretation of Poincaré's conventionalism is that it leads to a difficulty in saying why we should be, as Poincaré held, conventionalists about geometry in particular, rather than about other sciences. For, if our conventionalism derives from that fact that once terms occurring in axiom schemata are coordinated with reality, the schemata may be held to come what may then the same can be done for fields besides geometry, such as kinematics.

As Michael Friedman has argued, only an appeal to the priority of group theory in Poincaré's hierarchy of sciences can make sense of why he was a conventionalist about spatial geometry and not kinematics:

> Now Poincaré's conception of geometry is also very similar to the Kantian conception of geometry. For Poincaré, as for Kant, geometry is synthetic, because it is based, like arithmetic, on the possibility of indefinitely repeating particular operations: namely, group-theoretical operations constituting a Lie group of free motions. Moreover, geometry is viewed as the presupposition of all empirical physical theories: neither for Poincaré nor for Kant can geometry itself be either empirically confirmed or empirically disconfirmed. The difference, of course, is that Poincaré, in contrast to Kant, is acquainted with *alternative* geometries. Poincaré is acquainted, in particular, with the Helmholtz-Lie theorem, according to which geometry is constrained, but by no means *uniquely determined*, by the idea of a Lie group of free motions. It then follows for Poincaré because three alternative possibilities are still left open, that we have here—in this very special situation—a conventional choice or free stipulation (Friedman 1997, p. 312).

[11] Compare the "monotonous series of reasonings" characteristic of arithmetic and the "long memoirs" on the classification of groups. Poincaré's impatience with this sort of work, as an "intuitive mathematician," should not be read as a dismissal of its foundational importance. See (McLarty 1997) for more on Poincaré's view of rigor.

There is a long story to be told about the relationship between Poincaré's hierarchical view of science and his restricted conventionalist thesis. A full telling is beyond the scope of the present discussion, but it suffices to concur with Friedman that the revised interpretation is better situated than the established view to make sense of why Poincaré's conventionalism is restricted to geometry and does not tend toward more radical versions of confirmational holism based on the contention that *any* proposition may be held onto in the face of apparently contradictory evidence, as endorsed, for example, by Quine (Quine 1951). According to the established view of Poincaré's conventionalism, geometry is conventional because it is non-falsifiable. To block falsification one simply handles anomalies by altering the kinematics. However, the relationship between geometry and kinematics is symmetrical on this account. Should we choose to stubbornly adopt a theory in kinematics, we could resist falsification by making compensating changes to our geometry. On the revised interpretation, geometry is non-falsifiable because it is conventional, a choice of a group must be made before empirical science can proceed.

7 Defining Groups

I have argued that, as with arithmetic, Poincaré understood groups to be generated by potentially indefinite iteration of a generating element. In the case of discrete groups, this may be done by specifying generators and relations. In the case of continuous groups, we need infinitesimal generators. It may be objected that this approach, after all, leans on implicit definitions, given the formalist character of group presentations. However, even if it is granted that group theory in general rests on implicit definition, the definitions of the terms contained in the typical axioms of geometry would not be implicitly defined. In fact, there is good reason to think that Poincaré would want implicit definition to be restricted to the algebraic setting, given his comments on definition in his later works. In an essay contained in the collection *Dernières Pensées*, Poincaré distinguishes two cases of what he calls "definition by postulate," which may plausibly be considered to be implicit definitions. The case which is objectionable to the "pragmatist" and accepted by the "Cantorian" occurs when an object is defined implicitly by a postulate which relates it to an infinity of elements in a supposedly given domain. However, this form of definition is acceptable to the pragmatist when an element is defined by its relation to objects in a finite domain (Poincaré 1963, p. 69–71). Katherine Dunlop has pointed to these passages to argue that Poincaré cannot succeed in founding geometry on group theory in a way that does not run afoul of his restrictions on definition, because "a group is characterized by the existence of an identity element that bears a specified relation to every object in the group" (Dunlop 2016, p. 303). However, when a group is presented by a finite set of generators and their relations, only the relation of the generators to the identity must be stipulated, and the universal generalization to all elements can be proven by an inductive argument. On the other hand, axioms of geometry, taken as implicit definitions, would run afoul of this constraint on definitions; for, they aim to simultaneously define infinitely many points and lines by simultaneously stipulating relations to other objects in the same infinite set.

Furthermore, I would contend that Poincaré's account of indefinite repetition and continuity, both of which will be involved in constructions from infinitesimal generators, appeals only to a kind of intellectual intuition, and not, as Janet Folina has argued, to Kantian forms of sensible intuition (Folina 1992, p. 35–36, p. 134–135, p. 180–181). Poincaré's views on these matters are, however, subtle and possibly shifted in his later texts in ways that are favorable to Folina's reading. A competent and complete engagement with Folina's thorough treatment of Poincaré's views is beyond our present scope. If Folina's view is correct, then there is an ineliminable appeal to forms of sensibility that is required to adequately define the groups and sub-groups that we have been discussing. While this would undermine the position that Poincaré rejects forms of sensibility and founds mathematics entirely on forms of the understanding, which is the strongest view I am inclined to endorse, it leaves untouched my core contention that group theory is prior in conception to the axioms of geometry.

8 Poincaré and His Contemporaries

A recurring theme of this paper has been the comparison of Poincaré's and Hilbert's images of mathematics. According to the established view there is little difference between the two on geometry, with the primary difference coming in Poincaré's proto-intuitionist view of arithmetic and the irreducibility of arithmetic induction to logic.[12] According to the view espoused in this paper, Poincaré held that group theory was the foundation of both arithmetic and geometry. Arithmetic concerns the important special case: the infinite cyclic group, otherwise known as the integers. The groups that constitute the forms of the understanding, however, include continuous groups with the invariant sub-groups that constitute the points and lines about which axioms assert truths. In each case, the construction of these forms occurs through the mind's power to adjoin and combine operations.

Poincaré, to be sure, well appreciated the power of the model theoretic methods employed by Hilbert in his *Grundlagen der Geometrie*, and praised them in his review. Poincaré especially praised the usefulness of Hilbert's work for breaking down the bias of geometers toward thinking that there is a "general geometry," with Euclidian, Lobachevskian, and Riemannian geometry as special cases. Hilbert's early model theoretic approach effectively demonstrated that it was not only the parallel postulate that could be regarded as logically independent:

[12] A nice discussion of the relationship between Poincaré and full-fledged Brouwerian intuitionism can be found in Folina's book *Poincaré and the Philosophy of Mathematics* (Folina 1992, p. 73–74).

> But why, among all the axioms of geometry, should [the parallel postulate] be the only one which could be denied without offense to logic? Whence should it derive this privilege? There seems to be no good reason for this, and many other conceptions are possible (Poincaré 1903, p. 2).

In particular, Hilbert's approach to independence proofs demonstrates the existence of non-Pascalian and non-Desarguesian geometries. Throughout the review, Poincaré demonstrates a clear understanding of Hilbert's methods and an admiration for the results obtained, but does this indicate agreement with the view that the supposedly basic terms of geometry are defined contextually by implicit definitions?

I think that the answer is clearly "no," and this is shown by Poincaré's concluding remarks from the review. There, Poincaré again emphasizes the group theoretical standpoint. First, he notes that, according to Hilbert, "The objects which he calls points, straight lines, or planes become thus purely logical entities which it is impossible to represent to ourselves," but he does not endorse this view (Poincaré 1903, 22). Rather, he criticizes Hilbert for neglecting the group theoretic definition of these terms, and in particular for failing to note that the non-Pascalian geometries that Hilbert studies may be generated from geometric groups in a way that is analogous to the generation of metrical geometries. The passage culminates in the following criticism:

> Professor Hilbert seems rather to slur over these inter-relations; I do not know why. The logical point of view alone appears to interest him. Being given a sequence of propositions, he finds that all follow logically from the first. With the foundation of this first proposition, with its psychological origin, he does not concern himself. And even if we have, for example, three propositions A, B, C, and if it is logically possible, by starting with any one among them, to deduce the other two from it, it will be immaterial to him whether we regard A as an axiom, and derive B and C from it, or whether, on the contrary, we regard C as an axiom, and derive A and B from it. The axioms are postulated; we do not know where they come from; it is then as easy to postulate A as C (Poincaré 1903, p. 22)

If Poincaré held the view that in geometry axioms are arbitrary postulates that implicitly define geometric terms, then this criticism would make no sense. The criticism only makes sense if, as I have argued, Poincaré holds that there is a correct account of where the axioms come from, and this account, as should by now be manifestly evident, requires that the terms contained in the axioms of geometry have proper definitions. According to Poincaré, there are non-logical criteria for the correct ordering of propositions in mathematics.

While the Poincaré-Russell polemics have been commonly interpreted as paralleling the Hilbert-Frege correspondence, the orthogonality between Poincaré and Hilbert is in fact previsioned in his review of Russell. Coffa in particular, regards Poincaré as having rejected the thesis of "semantic atomism," according to which the terms contained in a sentence must have prior meanings that must be properly combined for the sentence to express a proposition (Coffa 1993, p. 131). If this is right, then there is indeed a parallel between Frege's insistence that the axioms of geometry must be assertions about objects that are given in intuition and Russell's supposed semantic atomism.[13] However, if Poincaré is rejecting semantic atomism, then there cannot be correct definitions of the terms "point" and "line," as the 1898 *Monist* article articulates, that imply the axioms, as described in *Science and Hypothesis*.[14]

In fact, Poincaré's agenda in *Des Fondaments de Geométrie*, the review of Russell's 1897 book *An Essay on the Foundations of Geometry*, is entirely consonant with his agenda in his review of Hilbert. Firstly, he wishes to disavow Russell of the view that there is a core geometry, based on the axioms of projective geometry, that is grounded in sensible intuition. This is precisely the sort of view that Poincaré praises Hilbert for disproving by demonstrating the logical possibility of non-Pascalian and non-Desarguesian geometries through model theoretical methods (though, again, Poincaré thought these methods were incomplete because they did not show the origin of these geometries in group theory). Russell's contention, in the 1890s, was that projective geometry rested on a Kantian foundation in sensible intuition and consisted of synthetic *a priori* truths, but that metrical geometry was empirical.

It is true that Poincaré savagely criticizes Russell's position that the basic terms of projective geometry refer to objects given in sensible intuition, by pointing out the vagueness and imprecision of Russell's definitions and axioms (Poincaré 1899, §2–4), but this only shows that Poincaré rejected the foundation of geometry in sensibility. As I have argued, his alternative was not a foundation of geometry in discursive concepts defined implicitly by the axioms in which they occur, but rather in group theory as a form of the understanding. Indeed, the group theoretical perspective is what Poincaré insists on, and not the rejection of semantic atomism, in presenting his alternative to Russell (Poincaré 1899, §19). Poincaré wants to disavow Russell of the view he praises Hilbert for dislodging, that there is a universal geometry (for Russell, projective geometry), but he wants to do so by convincing Russell of the priority in conception of the form of the group, which he criticizes Hilbert for neglecting.

There is a clear affinity between Poincaré's view and that of Klein and Lie, expressed in the Erlangen program. Indeed, there was interaction between Poincaré and Lie in 1882 when Lie was in Paris. Lie wrote to Klein that Poincaré held the concept of a group to be the fundamental concept for all of mathematics (Hawkins 2000, p. 182). Jeremy Gray notes that it is very likely that Poincaré's use of groups in his analysis of Fuchsian functions was independent of Klein's Erlangen program (Gray 2005, p. 551). It is not clear whether Klein or Lie ever indicated anything like Poincaré's philosophical view that the group is prior in conception to the terms

[13] I think that there are complications with the interpretation of Russell's views in the 1890s implicit in Coffa's commentary, but I limit myself here to correcting the record on Poincaré.

[14] See my discussion above in the section titled "Definitions in Disguise." Notably, the 1898 article was written prior to the publication in 1899 of Poincaré's review of Russell's book on the foundations of geometry. Poincaré's view of geometry as founded in groups theoretical forms of the understanding was motivated by his early mathematical work and held consistently through his writings.

occurring in the axioms of geometry. Poincaré's position is that group theory defines those terms, that the objects to which they refer are constituted by invariant sub-groups, but an alternative view would be that the objects of geometry are presented or constructed independently while groups afford a complete means of classification without necessarily constituting the objects that comprise the spaces thus classified. This view is consistent with a mathematical interest in the use of groups to classify geometries, but inconsistent with Poincaré's fully developed philosophical position.

9 Conclusion

This paper has been focused on correctly articulating Poincaré's views on the foundations of geometry. Accordingly, he held that geometry is founded on group theory, in such a way that the terms "point" and "line" receive proper definitions by reference to invariant sub-groups. This interpretation sheds light on Poincaré's general philosophy of mathematics. For, group theory, as the study of forms of the understanding, has the same foundation as arithmetic. While Poincaré has sometimes been read as having a disjointed view of mathematics, a kind of quasi-logicist view of geometry as grounded in axiomatic implicit definitions together with a quasi-intuitionistic view of arithmetic, we can now see that there is a deep unity in his vision of mathematics.

Should this matter to contemporary philosophers of mathematics, or is getting Poincaré's view right a matter of historical curiosity? I think that, for good reason, philosophers have been fascinated with the notion of mathematics as a science of structure, or logical form, which can be associated with the concepts that are articulated through axiomatic definitions. This has given rise to a very interesting literature in the metaphysics of mathematics, on whether structures must be understood as *ante rem* or *in re*, what the relationship between structures and sets might be, and many more questions. So it may seem that revisiting a neo-Kantian view from the turn of the 20th century holds limited interest. From this perspective, one may even find that my interpretation of Poincaré in fact diminishes his legacy as anticipating the structuralist turn, based on the axiomatic method, in his writings of the 1890s.

However, while the importance of the axiomatic point of view to the development of mathematics in the 20th century is undeniable, and the understanding of mathematics as advancing through the logical derivation of theorems from axioms, that is as a system of discursive reasoning, certainly has great philosophical importance, the focus on discursive reasoning limits our perspective on the nature of mathematical reasoning, on mathematical experience, and potentially on the very nature of the mind and intelligence. Poincaré's unified image of mathematics, derived from his own mathematical experience, beginning with his work on Fuchsian functions, which produced germinal contributions to mathematical knowledge, begins with an account of non-discursive reasoning. Mathematicians are masters of more than the rules of logic, and a complete account of mathematics will involve accounting for both discursive and non-discursive reasoning, as well as the relationship between the two. Correctly understood, Poincaré's account of the foundations of geometry offers a neglected contribution to this philosophical project.

References

P. A. Blanchette. Frege and Hilbert on consistency. *Journal of Philosophy*, 93(7):317–336, 1996.

O. Blumenthal. Lebensgeschichte. In *David Hilbert: Gesammelte Abhandlungen*, volume III, pages 388–429. Springer, 1935.

R. Carnap. *The Logical Structure of the World: And, Pseudoproblems in Philosophy*. Open Court classics. Open Court, 1969.

J. Coffa. *The Semantic Tradition from Kant to Carnap: To the Vienna Station*. Cambridge University Press, 1993.

W. Demopoulos. Frege, Hilbert, and the conceptual structure of model theory. *History and Philosophy of Logic*, pages 211–255, 1994.

K. Dunlop. Poincaré on the foundations of arithmetic and geometry. part 1: Against "dependence-hierarchy" interpretations. *Hopos: The Journal of the International Society for the History of Philosophy of Science*, 6(2):274–308, 2016.

J. Folina. *Poincaré and the Philosophy of Mathematics*. St. Martin's Press, 1992.

M. Friedman. Kant on the mathematical sciences. *Synthese*, 84:213–257, 1990.

M. Friedman. Poincaré's conventionalism and the logical positivists. In *Machine Discovery*, pages 299–314. Springer Netherlands, 1997.

J. Gray. Felix Klein's Erlangen program, 'Comparative considerations of recent geometrical researches'. In I. Grattan-Guinness, editor, *Landmark Writings in Western Mathematics 1640–1940*. Elsevier Science, 2005.

T. Hawkins. *The Emergence of the Theory of Lie Groups: An Essay in the History of Mathematics, 1869–1926*. Sources and Studies in the History of Mathematics and Physical Sciences. Springer Verlag, 2000.

G. Heinzmann. Poincaré on understanding mathematics. *Philosophia Scientiae*, 3(2):43–60, 1998.

G. Heinzmann and D. Stump. Henri Poincaré. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2014 edition, 2014.

C. McLarty. Poincaré: Mathematics & Logic & Intuition. *Philosophia Mathematica*, 5:97–115, 1997.

H. Poincaré. On the foundations of geometry. *The Monist*, 9(1):1–43, 1898.

H. Poincaré. Des fondements de la geométrie. *Revue de Métaphysique et de Morale*, 7:251–279, 1899.

H. Poincaré. Poincaré's review of Hilbert's "foundations of geometry.". *Bulletin of the American Mathematical Society*, 10(1):1–23, 1903.

H. Poincaré. *Science and Hypothesis*. The Walter Scott Publishing Co., 1905.

H. Poincaré. *Mathematics and Science: Last Essays*. Dover Publications, 1963.

H. Poincaré. Analyse de ses travaux scientifiques. *Proceedings of Symposia in Pure Mathematics*, 39(2):257–357, 1983.

H. Poincaré. *Papers on Fuchsian Functions*. Springer-Verlag, 1985.

H. Poincaré. *Three Supplements on Fuchsian Functions*. Akademie-Verlag, 1997.

W. Quine. Two dogmas of empiricism. *Philosophical Review*, 60(1):20–43, 1951.

M. Schlick and A. Blumberg. *General Theory of Knowledge*. Library of Exact Philosophy. Open Court Pub., 1974.

S. Shapiro. *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press, 2000.

J. Shipley. Frege on the foundation of geometry in intuition. *Journal for the History of Analytical Philosophy*, 3(6), 2015.

N. K. Smith. *Immanuel Kant's Critique of Pure Reason*. Macmillan, 1929.

J. Tappenden. Geometry and generality in Frege's philosophy of arithmetic. *Synthese*, 102(3): 319–361, 1995.

M. Wilson. Frege: The royal road from geometry. *Noûs*, 26(2):149–180, 1992.

Natural Deduction for Diagonal Operators

Fabio Lampert

Abstract We present a sound and complete Fitch-style natural deduction system for an **S5** modal logic containing an actuality operator, a diagonal necessity operator, and a diagonal possibility operator. The logic is two-dimensional, where we evaluate sentences with respect to both an actual world (first dimension) and a world of evaluation (second dimension). The diagonal necessity operator behaves as a quantifier over every point on the diagonal between actual worlds and worlds of evaluation, while the diagonal possibility quantifies over some point on the diagonal. Thus, they are just like the epistemic operators for apriority and its dual. We take this extension of Fitch's familiar derivation system to be a very natural one, since the new rules and labeled lines hereby introduced preserve the structure of Fitch's own rules for the modal case.

Keywords Modal logic • Natural deduction • Two-dimensional modal logic • Two-dimensional semantics • A priori • Actuality • Necessity

1 Introduction

One familiar doctrine in contemporary analytic philosophy says, roughly, that a sentence is a priori knowable if and only if it is true along the diagonal. Leaving aside several important qualifications and distinctions, this characterization of the a priori and others very similar to it have appeared in Stalnaker (1978), Evans (1979), Davies and Humberstone (1980), Kaplan (1989), Jackson (1998), Chalmers (1996, 2004), and others. These constitute the core of the research program identified as *two-dimensional semantics*.[1] In general, two-dimensional semanticists defend

[1] If we can say that these constitute a single research program at all. In any case, Chalmers (2004) provides a nice account of how exactly most of these differ.

F. Lampert (✉)
University of California, Davis, CA, USA
e-mail: fabiodcl@yahoo.com.br

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*, Proceedings of the Canadian Society for History and Philosophy of Mathematics La Société Canadienne d'Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_3

Fig. 1 A 2D-matrix for a diagonally true sentence φ

$$\begin{pmatrix} & w_0 & w_1 & w_2 & w_3 & \cdots \\ w_0 & \varphi & & & & \\ w_1 & & \varphi & & & \\ w_2 & & & \varphi & & \\ w_3 & & & & \varphi & \\ \vdots & & & & & \ddots \end{pmatrix}$$

an extension of the formal apparatus of possible worlds semantics in order to analyse the meaning of linguistic expressions. They claim that a sentence should be evaluated not only with respect to a single index or possible world, but also to a pair thereof, whereby different kinds of two-dimensional semantics will then claim distinct interpretations for the new indices. To say that a sentence is a priori knowable then comes down to it being true at every pair of indices where the first and second coordinate are identical, that is, the diagonal between pairs of indices/possible worlds (see Figure 1).

Extant presentations of the formalism involved in several versions of two-dimensionalism have been either semantic or, when proof theoretic, axiomatic in character.[2] For example, Davies and Humberstone (1980) present an axiomatization of a two-dimensional system that arguably captures some important a priori truths as a formal rendering of Evans' (1979) notion of *deep necessity*.[3] Evans distinguished between superficial and deep necessity in order to account for Kripke's (1980) examples of the contingent a priori. For Evans, superficial necessity or contingency is a matter of how a sentence behaves when embedded within the modal operators $\Box$ and $\Diamond$; by contrast, deep necessity and contingency have to do with the necessity or contingency of the content expressed by the sentence in question. By stipulating that 'Julius', for instance, names the inventor of the zip, the sentence (S) 'if anyone uniquely invented the zip, Julius invented the zip' is deeply necessary, for by understanding what is expressed by (S) we know it merely says that if anyone uniquely invented the zip, it was the inventor of the zip himself. And this, i.e. what is expressed by (S), is knowable a priori. Nonetheless, (S) is superficially contingent, for it is easily seen to be false at every possible world where the inventor of the zip differs from William Whitworth—the *actual* inventor of the zip. Davies and Humberstone's idea was to formalize these two notions of necessity (contingency) by adding to a modal language their 'fixedly actually' operator (see fn. 3) alongside the usual modal operators $\Box$ and $\Diamond$. In their system, formulas are

[2]The exception is Restall (2012), as mentioned below.

[3]Thereby relegating $\Box$ to *superficial necessity*. Strictly speaking, the axiomatic system in Davies and Humberstone (1980) does not contain a diagonal operator, but a 'fixedly' operator, $\mathscr{F}$, which alongside an actuality operator, $\mathscr{A}$, generates truth on the diagonal. It is with respect to $\mathscr{F}\mathscr{A}$ that deep necessity is defined. For more on the notions of deep and superficial necessity, see Evans (1979).

evaluated with respect to pairs of possible worlds, whence the logic is said to be two-dimensional. Each pair is constituted by a possible world and a world that is considered (intuitively) as the actual one. Thus, a sentence φ is fixedly actually true just in case it holds along the diagonal or, in other words, just in case φ is true at every possible world that is considered as the actual one.[4] Since the examples that are relevant to the distinction made by Evans between superficial and deep necessity (contingency) can be formalized within their two-dimensional system, Davies and Humberstone claimed to have a logic in which these two kinds of modalities are represented.

In addition to Davies and Humberstone's axiomatic system, there is a hypersequent system in Restall (2012) and an (equivalent) axiomatic system in Fritz (2013, 2014) designed to capture some of the core formal aspects of versions of two-dimensional semantics. These systems involve the usual modal operators, a primitive diagonal (apriority) operator, and an actuality operator. In particular, Fritz's system is motivated by the epistemic two-dimensionalism developed by Chalmers, where the indices of evaluation are intuitively interpreted as pairs constituted by epistemic scenarios and possible worlds, or epistemic and metaphysical possibilities, respectively. Since there is an epistemic possibility in which water is not H_2O (i.e. the XYZ-possibility), the sentence (S*) 'water is H_2O' is not a priori knowable, although it is necessary. More formally, we can say that there is a point along the diagonal where (S*) is false, but no point along the horizontal where (S*) is false (see, again, Figure 1). In contrast, there is no epistemic possibility where (S**) 'water is water' fails to hold, whence (S**) holds at every scenario-world pair along the diagonal. Thus, a sentence is a priori just in case it holds along the diagonal, and a posteriori otherwise.

The aim of this paper is to present a Natural Deduction system that exactly captures the behaviour of the diagonal operator and its interaction with the necessity operator in an **S5** modal logic with an added actuality operator, $\mathscr{A}$.[5] Modal systems containing $\mathscr{A}$ are by no means novelty. Some representatives are found as early as in Crossley and Humberstone (1977)—an axiomatic system—and Hazen (1978)—a natural deduction system. However, we know of no natural deduction system for two-dimensional modal logics. One of the virtues of the present system is that it is extremely easy to use for anyone already familiarized with traditional Fitch-style natural deduction systems for modal logics, thereby making the logical properties of the diagonal modalities, and some of the core ideas of two-dimensional semantics, transparent to a wider range of readers, in contrast with the (usually more cumbersome) axiomatic treatments.[6]

[4]Note that (S) holds at every world considered as actual, since in these worlds 'Julius' always picks out the *actual* inventor of the zip. If, by contrast, the actual world is held fixed, and we only consider different possible worlds relative to this world as the actual one, then there will be pairs at which (S) is false—precisely those in which the inventor of the zip differs from the actual one.

[5]We purposefully conflate use/mention when the context is clear to avoid cluttering the paper.

[6]In Lampert (forthcoming) we present several prefixed tableau systems for different two-dimensional modal logics.

2 The System $S5E_{2D}$

Definition 1 (Basic Language) Let $p, q, r \ldots$ be a countable list of members of the set AT of propositional letters. The language $\mathscr{L}_{\mathscr{A}\mathscr{D}}$ is defined recursively by the following grammar:

$$\varphi ::= AT \mid \neg\varphi \mid (\varphi \wedge \psi) \mid \Box\varphi \mid \mathscr{A}\varphi \mid \mathscr{D}\varphi$$

The other Boolean connectives are defined as usual, and we also define the duals $\Diamond := \neg\Box\neg$ and $\mathscr{C} := \neg\mathscr{D}\neg$.

Definition 2 (2D-Model) A *2D-model* is a tuple, $\mathscr{M} = \langle W, \mathscr{R}_\Box, \mathscr{R}_\mathscr{D}, V\rangle$, where

1. $W = Z \times Z$ for some set Z,
2. $\mathscr{R}_\Box \subseteq W \times W$, the $\Box$-accessibility relation, is the least relation such that for every $v, w, z \in Z$, $\langle w, v\rangle \mathscr{R}_\Box \langle z, v\rangle$,
3. $\mathscr{R}_\mathscr{D} \subseteq W \times W$, the $\mathscr{D}$-accessibility relation, is the least relation such that for every $v, w, z \in Z$, $\langle w, v\rangle \mathscr{R}_\mathscr{D} \langle z, z\rangle$, and
4. V is a function assigning sets $V(p)$ of pairs of possible worlds to each member of AT.

Definition 3 (Truth) We define 'φ is true at w with respect to v in $\mathscr{M}$', written $\mathscr{M}, \langle w, v\rangle \vDash \varphi$, by recursion on φ. Here w is the typical world of evaluation while v is the actual world under consideration. Intuitively, v might also be taken as an epistemic scenario or possibility. For a pair $\langle w, v\rangle \in W$,

$$\begin{array}{lll}
\mathscr{M}, \langle w, v\rangle \vDash p & \Longleftrightarrow & \langle w, v\rangle \in V(p); \\
\mathscr{M}, \langle w, v\rangle \vDash \neg\varphi & \Longleftrightarrow & \mathscr{M}, \langle w, v\rangle \nvDash \varphi \\
\mathscr{M}, \langle w, v\rangle \vDash (\varphi \wedge \psi) & \Longleftrightarrow & \mathscr{M}, \langle w, v\rangle \vDash \varphi \text{ and } \mathscr{M}, \langle w, v\rangle \vDash \psi; \\
\mathscr{M}, \langle w, v\rangle \vDash \Diamond\varphi & \Longleftrightarrow & \text{for some } z \in Z, \langle w, v\rangle \mathscr{R}_\Box \langle z, v\rangle \text{ and } \mathscr{M}, \langle z, v\rangle \vDash \varphi; \\
\mathscr{M}, \langle w, v\rangle \vDash \Box\varphi & \Longleftrightarrow & \text{for every } z \in Z \textit{ such that} \langle w, v\rangle \mathscr{R}_\Box \langle z, v\rangle, \mathscr{M}, \langle z, v\rangle \vDash \varphi; \\
\mathscr{M}, \langle w, v\rangle \vDash \mathscr{A}\varphi & \Longleftrightarrow & \mathscr{M}, \langle v, v\rangle \vDash \varphi; \\
\mathscr{M}, \langle w, v\rangle \vDash \mathscr{C}\varphi & \Longleftrightarrow & \text{for some } z \in Z, \langle w, v\rangle \mathscr{R}_\mathscr{D} \langle z, z\rangle \text{ and } \mathscr{M}, \langle z, z\rangle \vDash \varphi. \\
\mathscr{M}, \langle w, v\rangle \vDash \mathscr{D}\varphi & \Longleftrightarrow & \text{for every } z \in Z \textit{ such that} \langle w, v\rangle \mathscr{R}_\mathscr{D} \langle z, z\rangle, \mathscr{M}, \langle z, z\rangle \vDash \varphi.
\end{array}$$

Definition 4 (Truth, Validity, and Consequence) A sentence, φ, is *true in a 2D-model* $\mathscr{M} = \langle W, \mathscr{R}_\Box, \mathscr{R}_\mathscr{D}, V\rangle$, written $\mathscr{M} \vDash \varphi$, if and only if $\mathscr{M}, \langle w, w\rangle \vDash \varphi$ for every $\langle w, w\rangle \in W$; a sentence φ is *diagonally valid* in a class $\mathscr{C}$ of models if and only if it is true in every model in $\mathscr{C}$; and a sentence φ is a *logical consequence* of a set of sentences Γ if and only if for all models $\mathscr{M} = \langle W, \mathscr{R}_\Box, \mathscr{R}_\mathscr{D}, V\rangle$ and $\langle w, w\rangle \in W$, if $\mathscr{M}, \langle w, w\rangle \vDash \gamma$ for all $\gamma \in \Gamma$, then $\mathscr{M}, \langle w, w\rangle \vDash \varphi$.

3 Natural Deduction System for $\mathbf{S5E}_{2D}$

In Fitch's (1952) original presentation for modal logic a natural deduction system is enriched with $\Box$-labeled lines to denote the introduction of a possible world in a derivation. In other words, $\Box$-labeled lines are world-tags, for arbitrary possible worlds. This means that the formulas occurring in a derivation within their scope should be evaluated at the possible world introduced by the respective $\Box$-labeled line. The new subproofs introduced by $\Box$-labeled lines are called *strict* or just $\Box$-*subproofs*. In one-dimensional modal logic we use $\Box$-labeled lines in cases where we need to eliminate or introduce the modal operators $\Box$ and $\Diamond$. For instance, we can prove $\Diamond p \supset \Box \Diamond p$ in **S5** by the following derivation:

1	$\Diamond p$	P
2	$\Box \mid \Diamond p$	Mod Reit, 1
3	$\Box\Diamond p$	$\Box$I, 2
4	$\Diamond p \supset \Box\Diamond p$	$\supset$I, 1–4

By contrast, if we want to evaluate formulas with respect to a pair of worlds, as our two-dimensional framework requires, we should expect the system to be a bit more complicated, for our Fitch-lines must dictate that the formulas are true at a world *with respect to a certain world considered as actual*, or simply that the formula is true at a pair of possible worlds. Thus, there is a further dimension present in what our labeled lines ought to convey in the proof system. Our Fitch-lines, as a result, will comprise a $\Box$-labeled line, an actuality labeled line denoted by $\mathscr{A}$, and a diagonal labeled line denoted by δ, the latter introduced specifically for diagonal operators.

In what follows we show the Fitch rules necessary for $\mathbf{S5}\mathrm{E}_{2D}$. For ease of presentation, rather than listing all of them at once, first we recall the rules for **S5**, then we show the modifications needed to accommodate the actuality operator, and only after that we add the rules for the entire two-dimensional system. Thus, in effect, we present natural deduction systems for three different logics ordered by inclusion: **S5**, $\mathbf{S5}\mathscr{A}$, and $\mathbf{S5}\mathrm{E}_{2D}$.

3.1 S5

In **S5** the rules for the Boolean connectives are just the standard Fitch rules for propositional logic that one finds in any logic textbook. The rules governing the modal operators are the following:

(□I) □ Introduction

$$
\begin{array}{l|l|l}
\vdots & \Box & \vdots \\
i & & \varphi \\
\hline
j & \Box\varphi & \qquad \Box\mathrm{I},\ i
\end{array}
$$

(□E) □ Elimination

$$
\begin{array}{l|ll}
i & \Box\varphi & \\
\vdots & \vdots & \\
k & \varphi & \qquad \Box\mathrm{E},\ i
\end{array}
\qquad\qquad
\begin{array}{l|l|l}
i & \Box\varphi & \\
\vdots & \Box & \vdots \\
k & & \varphi \qquad \Box\mathrm{E},\ i
\end{array}
$$

(◇I) ◇ Introduction

$$
\begin{array}{l|l|l}
\vdots & \Box & \vdots \\
i & & \varphi \\
\hline
j & \Diamond\varphi & \qquad \Diamond\mathrm{I},\ i
\end{array}
\qquad\qquad
\begin{array}{l|ll}
\vdots & \vdots & \\
i & \varphi & \\
j & \Diamond\varphi & \qquad \Diamond\mathrm{I},\ i
\end{array}
$$

(◇E) ◇ Elimination

$$
\begin{array}{l|l|l}
i & \Diamond\varphi & \\
j & \Box & \varphi \\
\cline{3-3}
\vdots & & \vdots \\
k & & \psi \\
\hline
l & \Diamond\psi & \qquad \Diamond\mathrm{E},\ i,\ j\text{–}k
\end{array}
$$

(MN) Modal Negation

$$
\begin{array}{l|ll}
i & \neg\Box\varphi & \\
\vdots & \vdots & \\
k & \Diamond\neg\varphi & \qquad \mathrm{MN},\ i
\end{array}
\qquad\qquad
\begin{array}{l|ll}
i & \Diamond\neg\varphi & \\
\vdots & \vdots & \\
k & \neg\Box\varphi & \qquad \mathrm{MN},\ i
\end{array}
$$

$$
\begin{array}{l|ll}
i & \neg\Diamond\varphi & \\
\vdots & \vdots & \\
k & \Box\neg\varphi & \qquad \mathrm{MN},\ i
\end{array}
\qquad\qquad
\begin{array}{l|ll}
i & \Box\neg\varphi & \\
\vdots & \vdots & \\
k & \neg\Diamond\varphi & \qquad \mathrm{MN},\ i
\end{array}
$$

3.2 S5$\mathscr{A}$

The derivation system for the actuality operator is essentially the same as the one in Hazen (1978). Let an $\mathscr{A}$-labeled line denote a new kind of strict subproof, which we might also call *$\mathscr{A}$-subproof*. Moreover, we now say that the main proof is also counted as an *$\mathscr{A}$-subproof*. Now add the following rules to the ones presented above for **S5**:

($\mathscr{A}$I) Actuality Introduction

$$\begin{array}{l|l|l l}
\vdots & \mathscr{A} & \vdots & \\
i & & \varphi & \\
j & \multicolumn{2}{l}{\mathscr{A}\varphi} & \mathscr{A}\text{I}, i
\end{array}$$

($\mathscr{A}$E) Actuality Elimination

$$\begin{array}{l|l|l l}
i & \multicolumn{2}{l}{\mathscr{A}\varphi} & \\
j & \mathscr{A} & \varphi & \\
\cline{3-3}
\vdots & & \vdots & \\
k & & \psi & \\
l & \multicolumn{2}{l}{\mathscr{A}\psi} & \mathscr{A}\text{E}, i, j\text{–}k
\end{array}$$

(act-I) act-I (only in $\mathscr{A}$-subproofs)

$$\begin{array}{l|l l}
a & \varphi & \\
\vdots & \vdots & \\
b & \mathscr{A}\varphi & \text{act-I}, a
\end{array}$$

(act-E) act-E (only in $\mathscr{A}$-subproofs)

$$\begin{array}{l|l l}
a & \mathscr{A}\varphi & \\
\vdots & \vdots & \\
b & \varphi & \text{act-E}, a
\end{array}$$

(N$\mathscr{A}$) Negated Actuality

$$\begin{array}{l|l l}
i & \neg\mathscr{A}\varphi & \\
\vdots & \vdots & \\
k & \mathscr{A}\neg\varphi & \text{N}\mathscr{A}, i
\end{array}
\qquad
\begin{array}{l|l l}
i & \mathscr{A}\neg\varphi & \\
\vdots & \vdots & \\
k & \neg\mathscr{A}\varphi & \text{N}\mathscr{A}, i
\end{array}$$

The rules of reiteration are essentially the same as for □-subproofs, i.e. $\mathscr{A}$-formulas can be reiterated into □-subproofs and vice versa. Hazen (1978, especially pp. 618–619) sketches the completeness theorem for the system **S5**$\mathscr{A}$.

3.3 $S5E_{2D}$

For the system **S5**E_{2D} we have all the rules presented above, besides small restrictions that we shall mention in due course. Next we introduce the rules for diagonal operators, which turn out to be very similar to the usual modal rules except for the addition of δ-subproofs.

($\mathscr{D}$I) Diagonal Necessity Introduction

$$\begin{array}{r|l|l l}
\vdots & \delta & \vdots & \\
i & & \varphi & \\
j & \mathscr{D}\varphi & & \mathscr{D}\text{I}, i
\end{array}$$

($\mathscr{D}$E) Diagonal Necessity Elimination

$$\begin{array}{r|l l}
i & \mathscr{D}\varphi & \\
\vdots & \vdots & \\
k & \varphi & \mathscr{D}\text{E}, i
\end{array}
\qquad
\begin{array}{r|l|l l}
i & \mathscr{D}\varphi & & \\
\vdots & \delta & \vdots & \\
k & & \varphi & \mathscr{D}\text{E}, i
\end{array}$$

($\mathscr{C}$I) Diagonal Possibility Introduction

$$\begin{array}{r|l|l l}
\vdots & \delta & \vdots & \\
i & & \varphi & \\
j & \mathscr{C}\varphi & & \mathscr{C}\text{I}, i
\end{array}
\qquad
\begin{array}{r|l l}
\vdots & \vdots & \\
i & \varphi & \\
j & \mathscr{C}\varphi & \mathscr{C}\text{I}, i
\end{array}$$

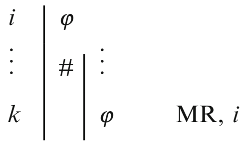
($\mathscr{C}$E) Diagonal Possibility Elimination

$$\begin{array}{r|l|l l}
i & \mathscr{C}\varphi & & \\
j & \delta & \underline{\varphi} & \\
\vdots & & \vdots & \\
k & & \psi & \\
l & \mathscr{C}\psi & & \mathscr{C}\text{E}, i, j\text{–}k
\end{array}$$

(DN) Diagonal Negation

$$\begin{array}{r|l l}
i & \neg\mathscr{D}\varphi & \\
\vdots & \vdots & \\
k & \mathscr{C}\neg\varphi & \text{DN}, i
\end{array}
\qquad
\begin{array}{r|l l}
i & \mathscr{C}\neg\varphi & \\
\vdots & \vdots & \\
k & \neg\mathscr{D}\varphi & \text{DN}, i
\end{array}$$

Fig. 2 Modal Reiteration Rule



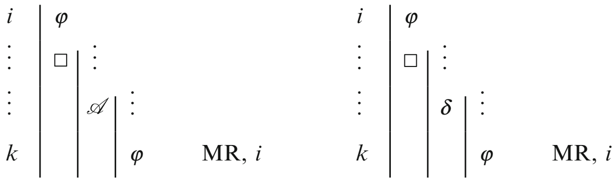
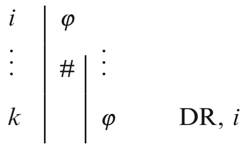
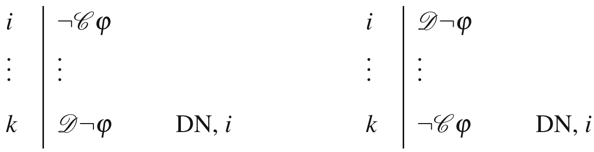


Fig. 3 Good and bad uses of modal reiteration

Fig. 4 Diagonal reiteration rule





Concerning the restrictions mentioned above, the rules for the Boolean connectives remain the same, the only exception being the *reductio ad absurdum* rules, for we allow the relevant contradiction to occur within the scope of any $\Box$, $\mathscr{A}$, or δ-subproof. More importantly, $\Box$, $\Diamond$, and $\mathscr{A}$-formulas can still be reiterated within $\Box$ and $\mathscr{A}$-subproofs, but they cannot be reiterated across δ-subproofs, thus, where $\# \in \{\Box, \mathscr{A}\}$, and φ is either a $\Box$, $\Diamond$, or $\mathscr{A}$-formula (Figure 2).

For example, the following instance of modal reiteration in the derivation on the left side is permissible, but the one on the right side is not (Figure 3):

The rule of *diagonal reiteration*, by contrast, can be used unrestrictedly in a derivation. Thus, in the derivation on the right side of Figure 3, the rule of diagonal reiteration (DR), rather than modal reiteration (MR), would allow the introduction of φ, more generally, where $\# \in \{\Box, \mathscr{A}, \delta\}$, and φ is any formula beginning with a diagonal operator (Figure 4).

Finally, the (act-I) and (act-E) rules are also allowed within δ-subproofs, while the other rules remain unchanged. In what follows we present some examples of derivations in **S5**E_{2D}, commenting (when necessary) on the usage of the rules just presented.

$$\vdash \mathscr{D}p \supset \Box\mathscr{D}p$$

1	$\quad \vert\ \mathscr{D}p$	P
2	$\quad \vert\ \Box \vert\ \mathscr{D}p$	DR, 1
3	$\quad \vert\ \Box\mathscr{D}p$	$\Box$I, 2
4	$\mathscr{D}p \supset \Box\mathscr{D}p$	$\supset$I, 1–3

In this case, $\mathscr{D}p$ was legitimately reiterated over a $\Box$-subproof since the reiteration rule for diagonal operators is unrestricted, while line 3 was obtained by usual application of $\Box$ introduction. The following exemplifies a distinction between the two operators $\Box$ and $\mathscr{D}$, especially in how they interact with the actuality operator:

$$\vdash \mathscr{D}(\mathscr{A}p \supset p)$$

1	$\delta \vert\ \vert\ \mathscr{A}p$	P
2	$\quad \vert\ \vert\ p$	act-E, 1
3	$\quad \vert\ \mathscr{A}p \supset p$	$\supset$I, 1–2
4	$\mathscr{D}(\mathscr{A}p \supset p)$	$\mathscr{D}$I, 1–3

$$\nvdash \Box(\mathscr{A}p \supset p)$$

1	$\Box \vert\ \vert\ \mathscr{A}p$	P
2	$\quad \vert\ \vert\ p$	<u>act-E</u> , 1
3	$\quad \vert\ \mathscr{A}p \supset p$	$\supset$I, 1–2
4	$\Box(\mathscr{A}p \supset p)$	$\Box$I, 1–3

The problem with the derivation on the right is that (act-E) is not allowed unless it is applied within $\mathscr{A}$ and δ-subproofs, as illustrated by the derivation on the left. In general, derivations containing only diagonal or the basic modal operators $\Box$ and $\Diamond$ will be analogous, but there are subtle differences in their interaction with $\mathscr{A}$.

4 Completeness for S5E$_{2D}$

The derivation system defined above is clearly sound with respect to the two-dimensional semantics presented in §2. Completeness, however, is not so obvious. We could establish it by a traditional Henkin-style argument extended to take care of pairs of possible worlds; however, it suffices to prove the axioms (by means of schematic derivations) of the system **2D** presented in Fritz (2013). In effect, Fritz defines two systems, to wit, **2D** and **2Dg**. The former is based on a semantics just like the one defined in §2, whereas the latter characterizes truth in a model and the consequence relation in a general way, that is, with respect to *every* point in a 2D-model: a sentence, φ, is *generally true in a 2D-model* if and only if $\mathscr{M}, \langle w, v\rangle \vDash \varphi$ for every $\langle w, v\rangle \in W$; a sentence φ is *generally valid* in a class $\mathscr{C}$ of models if and only if it is generally true in every model in $\mathscr{C}$; and a sentence φ is a *general logical consequence* of a set of sentences Γ if and only if for all models $\mathscr{M} = \langle W, \mathscr{R}_\Box, \mathscr{R}_\mathscr{D}, V\rangle$ and $\langle w, v\rangle \in W$, if $\mathscr{M}, \langle w, v\rangle \vDash \gamma$ for all $\gamma \in \Gamma$, then $\mathscr{M}, \langle w, v\rangle \vDash \varphi$. While $\mathscr{A}\varphi \supset \varphi$ and $\mathscr{D}\varphi \supset \varphi$ are valid in **2D**, for instance, they are not valid in **2Dg**, where truth, consequence, and validity are defined generally.[7] But this distinction between general and diagonal validity can nonetheless be accommodated in our derivation system, whereby the latter is complete with respect to either notion of validity. The axioms[8] defining **2Dg** comprise the following *actuality base* from Crossley and Humberstone's system **S5**$\mathscr{A}$,[9] besides the usual **S5** axioms:

$$(\mathscr{A}1)\ \mathscr{A}(\mathscr{A}\varphi \supset \varphi). \qquad (\mathscr{A}3)\ \mathscr{A}\varphi \equiv \neg\mathscr{A}\neg\varphi.$$
$$(\mathscr{A}2)\ \mathscr{A}(\varphi \supset \psi) \supset (\mathscr{A}\varphi \supset \mathscr{A}\psi). \quad (\mathscr{A}4)\ \mathscr{A}\varphi \supset \Box\mathscr{A}\varphi.$$

Moreover, the following axioms for diagonal operators are added by Fritz, with the (notational) difference that he uses A rather than $\mathscr{D}$, and @ rather than $\mathscr{A}$:

$$(4_\mathscr{D})\ \mathscr{D}\varphi \supset \mathscr{D}\mathscr{D}\varphi. \ (\mathbf{I}_3)\ \mathscr{D}\varphi \supset \mathscr{A}\varphi.$$
$$(5_\mathscr{D})\ \mathscr{C}\varphi \supset \mathscr{D}\mathscr{C}\varphi. \ (\mathbf{I}_4)\ \mathscr{D}(\mathscr{A}\varphi \supset \varphi).$$

If truth and validity are defined generally, as in **2Dg**, then the main proof of a derivation is not an $\mathscr{A}$-subproof anymore. (Note that this imposes a restriction on the derivation rules (act-I) and (act-E), while the other rules for the actuality operator remain sound.) Furthermore, $\mathscr{D}$ is not reflexive in **2Dg**, whence the reflexive rules for ($\mathscr{D}$E) and ($\mathscr{C}$I) must also be dropped in this case.[10] All the other rules are sound with respect to general validity, and the axioms above can be easily derived. On the other hand, by defining truth and validity as in definition 2.4, we may add the following axioms to the ones above:

$$(\mathscr{A}5)\ \mathscr{A}\varphi \supset \varphi. \ (\mathbf{I}_T)\ \mathscr{D}\varphi \supset \varphi.$$

[7] For a discussion on different accounts of validity in two-dimensional logics, see Lampert (forthcoming).

[8] Strictly, axiom-schemata.

[9] Fritz does not list $\mathscr{A}1$ between the axioms, but this appears in Crossley and Humberstone's original formulation (1977, p. 14).

[10] These are the rules where no new scope line is opened in the elimination of $\mathscr{D}$ and the introduction of $\mathscr{C}$, respectively.

In fact, by just substituting ($\mathscr{A}1$) by ($\mathscr{A}5$) and adding ($\mathbf{I}_T$) we get a complete axiomatization for **2D**. The derivations of these two axioms in $\mathbf{S5E}_{2D}$ are obvious and all the rules presented in §3 are sound with respect to diagonal validity.

In the following we exhibit schematic derivations for some of the axioms as illustration. The axioms governing the actuality operator are derived in Hazen (1978) in the midst of his proof that $\mathscr{A}$ is eliminable when added to the basic modal language. So we are left with the axioms for the diagonal operators, which we derive below. (A schematic derivation for $\mathbf{I}_4$ is just as illustrated in §4.)

$$\vdash \mathscr{D}\varphi \supset \mathscr{D}\mathscr{D}\varphi$$

1	$\quad \mid \underline{\mathscr{D}\varphi}$	P
2	$\quad \mid \delta \mid \mathscr{D}\varphi$	DR, 1
3	$\quad \mid \mathscr{D}\mathscr{D}\varphi$	$\mathscr{D}$I, 2
4	$\mathscr{D}\varphi \supset \mathscr{D}\mathscr{D}\varphi$	⊃I, 1–3

$$\vdash \mathscr{C}\varphi \supset \mathscr{D}\mathscr{C}\varphi$$

1	$\quad \mid \underline{\mathscr{C}\varphi}$	P
2	$\quad \mid \delta \mid \mathscr{C}\varphi$	DR, 1
3	$\quad \mid \mathscr{D}\mathscr{C}\varphi$	$\mathscr{D}$I, 2
4	$\mathscr{C}\varphi \supset \mathscr{D}\mathscr{C}\varphi$	⊃I, 1–3

$$\vdash \mathscr{D}\varphi \supset \mathscr{A}\varphi$$

1	$\quad \mid \underline{\mathscr{D}\varphi}$	P
2	$\quad \mid \mathscr{A} \mid \mathscr{D}\varphi$	DR, 1
3	$\quad \mid \quad \mid \varphi$	$\mathscr{D}$E, 2
4	$\quad \mid \mathscr{A}\varphi$	$\mathscr{A}$I, 3
5	$\mathscr{D}\varphi \supset \mathscr{A}\varphi$	⊃I, 1–4

5 Concluding Remarks

We have shown a sound and complete Fitch-style natural deduction system for an **S5** modal logic endowed with diagonal operators and an actuality operator in a two-dimensional semantics. The derivation rules presented here are analogous to Fitch's own modal rules for the one-dimensional case, thereby preserving familiar features of this kind of natural deduction system. The procedure is somewhat straightforward once new labeled lines such as $\mathscr{A}$ and δ are introduced alongside the usual $\Box$-labeled line. This is enough to make it possible to evaluate formulas with respect to a pair of worlds as dictated by the two-dimensional character of the system.

Acknowledgements Thanks to Alexander Kocurek, Andrew Parisi, Shawn Standefer, the UC Davis Llemma group, and two anonymous reviewers for comments on earlier drafts that greatly improved this paper.

References

Chalmers, D (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, D (2004) Epistemic Two-Dimensional Semantics. *Philosophical Studies*, 118, 153–226.

Crossley, JN and Humberstone, L (1977) The logic of "actually". *Reports on Mathematical Logic* 8, 11–29.

Davies, M and Humberstone, L (1980) Two notions of necessity. *Philosophical Studies*, 38(1), 1–30.

Evans, G (1979) Reference and Contingency. *The Monist*, 62, 161–189.

Fitch, FB (1952) *Symbolic Logic: an Introduction*. New York: Ronald Press, Co.

Fritz, P (2013) A logic for epistemic two-dimensional semantics. *Synthese*, 190, 1753–1770.

Fritz, P (2014) What is the correct logic of necessity, actuality, and apriority? *The Review of Symbolic Logic*, 7(3), 385–414.

Hazen, AP (1978) The eliminability of the actuality operator in propositional modal logic. *Notre Dame Journal of Formal Logic*, 4, 617–622.

Jackson F (1998) *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford University Press.

Kaplan, D (1989) Demonstratives. In Almog J, Perry J and Wettstein H (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford: Oxford University Press.

Kripke, S (1980) *Naming and necessity*. Cambridge, MA: Harvard University Press. Lampert, F (forthcoming) Actuality, tableaux, and two-dimensional modal logics. *Erkenntnis*.

Restall, G (2012) A cut-free sequent system for two-dimensional modal logic, and why it matters. *Annals of Pure and Applied Logic*, 163, 1611–1623.

Stalnaker, RC (1978) Assertion. In Cole P (Ed.), *Syntax and Semantics: Pragmatics* (Vol. 9, pp. 315–339). New York: Academic Press.

What is a Symbol?

Valerie Allen

Abstract Focusing my discussion on the sixteenth and seventeenth centuries, I argue that the symbolic notation under development at the time reveals connections with rhetorical and poetic aesthetics. In the first section, I show how the mathematical strategies that notation facilitated rely on prudential rhetoric's sense of the opportune moment, also known as *kairos* or decorum. In the second, I show how the necessary balance within notation between compression and lucidity similarly relies on an aesthetic judgment that is essentially prudential. In the last section, I show how notation's ability to accommodate disjunctive values within a general symbol corresponds to similar capabilities in the poetic symbol. Through etymological analysis of the polysemy of its terminology, I demonstrate the philological orientation of early modern algebra.

Keywords Symbol • Algebra • Notation • Aesthetics

1 Introduction

By "symbol" in this paper I refer to "symbolic notation," and there is long established precedent to do so, as Joseph Moxon's textbook from 1700 demonstrates: "*Symboles*, are Letters used for Numbers in Algebra" (Moxon 1700, p. 167). The quotation appears in the *Oxford English Dictionary* as one of the earliest attested usages for the definition of symbol as "a written character or mark used to represent something; a letter, figure, or sign conventionally standing for some object, process, etc." (Oxford English Dictionary Online 2017, symbol n.1, 3). Changing my essential question, however, to "What is Symbolic Notation?" subtly narrows the discussion by eliminating any seeming connection between mathematics and the aesthetics of rhetoric or poetry. Yet, as I will argue, those connections persist even when delimiting discussion to notation *stricto sensu*. Although notation

V. Allen (✉)
John Jay College CUNY, 524 West 59th Street #7.63.04NB, New York, NY, 10019, USA
e-mail: vallen@jjay.cuny.edu

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*, Proceedings of the Canadian Society for History and Philosophy of Mathematics La Société Canadienne d'Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_4

generally avoids what Leibniz calls the "countless ambiguities" of ordinary language (Bochenski 1961, pp. 274–275), an etymological examination of the English word "notation" and its associated terms reveals semantic entanglements with rhetoric that call into question the referential purity of mathematical terminology.

The correspondence itself between early modern mathematics and rhetoric is by now well established. Giovanna Cifoletti argues that "the reform of mathematics and science in the sixteenth century was first conceived of as a reform of the [rhetorical] 'art of thinking'" (Cifoletti 2006, p. 376). For her, the origins of the new algebra were "essentially philological" (Cifoletti 2006, p. 388), while for G.E.R. Lloyd, mathematical *apodeixis* developed in part from rhetorical *epideixis* (Lloyd 1990, pp. 77–78). These studies trace the shared conceptual architecture of rhetoric and mathematics. My own analysis builds on that recently made by Travis Williams, who argues that mathematical notation is "enargetic," that is, that in naming its objects, it renders them visually vivid just as does rhetorical *enargeia* (Williams 2016). A general term in classical rhetoric, *enargeia* refers to descriptions so evocative that they appear to convert the reader or listener into an eyewitness. The orator Lysias allegedly could deliver his material so graphically that the listener felt he could "see the actions which are being described going on and that he is meeting face-to-face the characters in the orator's story" (Lanham 1991, p. 64). In so achieving that effect, *enargeia* blurs fundamental category distinctions between tell and show, diegesis and mimesis, description and naming, mention and use, even rhetoric and dialectic. Rhetoric aims to persuade by any available means rather than to argue by means of logic, yet *enargeia* causes one to see what one can only hear about, consequently to know that about which one only has an opinion.

The claim that notation describes even as it names is consequential because it suggests that notation constructs rather than merely labels mathematical ideas in the same way that rhetorical expression forms thought rather than supervenes a preexistent idea. A growing body of recent work already claims notation to be less the mere bearer of content than the means by which new knowledge or content is generated (Heeffer 2009; Heeffer and Van Dyck 2010; Rotman 2000; Serfati 2010). Major insights from semiotics, cognition theory, cognitive science, and historical epistemology together demonstrate the cognitive agency of notation (Heeffer 2015). Drawing in the following three sections from prudential rhetoric, aesthetic theory, and etymological analysis, I textualize early modern symbolic notation, thereby arguing for its creative and mediatory function.

2 "Rhetorical" Algebra

The term conventionally refers to mathematical work that is expressed in ordinary language rather than symbols. By this definition, rhetorical algebra has to be notation-free (Heeffer 2009). I argue, however, for an underlying kinship between the methods of rhetorical expression and those of calculations performed in algebraic notation, by consequence claiming that algebra can be fully notational and at the same time deeply rhetorical.

Algebraic eloquence starts with the equation, which makes possible the composition of meaningful sentences, such as “x is equal to y,” because, in grammatical terms, “is” serves as the copula between subject and complement. As the sequential logic of the medieval trivium demonstrates, grammar underlies rhetoric. At the heart of the equation lies the equals sign, the notation for which was introduced by Robert Recorde in 1557, although his parallel lines, originally about one cm long (depending on line length), have since been standardized and shortened to dash-length. Although his notation was neither immediately adopted as universal standard nor recognized at the time as an important advance in symbolic representation (Cajori 1928, pp. 297–298), the confusing variety at the time of various symbols for equality indexes the importance of his innovation. When Recorde introduced the notation, he valued it merely as a labor-saving device. “And to avoid the tedious repetition of these words ‘is equal to’ I will set as I do often in work use, a pair of parallels” (Recorde 2010, p. 242). He could just have easily written out in full “is equal to” (or Latin *æquatur*). But in retrospect, we appreciate the significance of the change. Michel Serfati observes how the equals sign “made it impossible to maintain the syntax of natural language in the text. In effect, a new symbolism expressing ideal interchangeability succeeded the predicate structure of rhetorical expression” (Serfati 2010, p. 108). (Gottlob Frege also noted how mathematical logic unyoked itself from the subject—predicate structure of traditional grammar (Frege 1970, p. 7).) There is reason to think that Thomas Harriot, who used Recorde’s notation, understood its symbolic force in a way its inventor did not (Tanner 1962, p. 166). Recorde’s example at the point when he introduces the equals sign, namely, “14 root plus 15 is equal to 71” (which we would write as $14x+15=71$), possesses “$14x+15$” as its subject in terms of traditional grammar and “is equal to 71” as its predicate. In terms of symbolic notation, the replacement of the words “is equal to” by the sign “=” flattens the ontological hierarchy of the sentence and bestows equal explanatory value on the phrases “$14x+15$” and “71.” Thus flattened, the statement now runs in either direction, where traditional grammar allows us to move only from left to right, from subject to predicate. In *Alice’s Adventures in Wonderland*, Lewis Carroll draws attention to this distinction between the grammars of ordinary and symbolic language when the Mad Hatter objects to Alice’s claim that meaning what she says is the same thing as saying what she means (Carroll 2013, p. 53).

> “Not the same thing a bit!” said the Hatter. “You might just as well say that ‘I see what I eat’ is the same thing as ‘I eat what I see’!”

By commuting in either direction, the equation also undoes the logical antecedence of the grammatical subject to its predicate. Prior to (and even after) Recorde’s innovation, operations on a sum such as “14 root plus 15 is equal to 71” comprised a series of successive acts during which the “equation” was out of kilter and not properly an equation at all. Herein lies the difference between, on the one hand, moving 15 from the subject (leaving 14 root) to the predicate (leaving 56) and, on the other, adding negative 15 to both sides of the equation in a single operation. What the equals symbol enables is an idealization of the equality relation, and hence the creation of a conceptual object that by definition is always in balance, despite the fact that while we manipulate equations we do so in a succession of operations distributed across real time. Notation invites us to separate the conceptual from the temporal, experiential, and mechanical domains, perhaps making it all the harder to think of the ideas it represents as being constituted in and by language.

In the equation proper, then, the terms of the mathematical sentence interchange with each other, something that could not happen in traditional subject—predicate grammar. Temporarily substituting a simpler term for a compound concept (as in u-substitution in calculus) "remained an inconceivable operation for the medieval rhetorical writing of mathematics" (Serfati 2010, p. 117). Yet even as notation appears to distance mathematical reasoning from verbal discourse, its kinship with rhetoric emerges in the following ways.

Symbolic notation enables the solution of mathematical problems by using strategies that are in essence rhetorical. It goes almost without saying that a solution often depends on a trivial manipulation that nonetheless generates non-trivial results, as, for example, when we multiply $\lim x\, 0[(1-\cos x)/x]$ by its conjugate to determine the limit at 0. The value of the expression does not alter and the dilated language adds nothing but tautology and circumlocution yet it changes the form in such a way that new meaning emerges. Rhetorical manipulation of language, especially in the domain of forensic oratory, achieves similar ends. Rhetorical tropes subtly "turn" (as the etymology of "trope" implies) the significance of words while rhetorical figures change the form of words in order to emphasize them (Lanham 1991, p. 155). Both rhetorician and mathematician alike recognize the formal rules of the game by which the lexical and notational manipulations meaningfully inflect the argument.

In the five parts of classical rhetoric—invention, arrangement, style, memory, and delivery—were two techniques of style (Lat. *elocutio*, Gr. *lexis*) so fundamental that they affected the text's conceptual arrangement (Lat. *dispositio*, Gr. *taxis*) because of how substantively they manipulated content (Baltzell 1967; Faral 1924, p. 61). Amplification and abbreviation comprise categories of stylistic device rather than individual devices, and, of the two, amplification is generally the more important for early modern rhetoric because it is associated with grandeur and the high style. The importance of amplification or "surplusage" is evident throughout George Puttenham's treatise on poetic style, published as Thomas Harriot was writing his mathematical papers. One figure "fit for amplification," antitheton, restates every idea in the negative (a strategy well known to mathematicians): "Ready to *ioy* your gaines, your losses to bemone, ... Who onely bred your blisse, and neuer causd your care" (Puttenham 1589, p. 175).

Less cherished than the copiousness of amplification, rhetorical abbreviation nonetheless plays a valuable part in directing thought. Indeed, one of its main devices declares its intention in its very name: in Greek, *emphasis*, in Latin, *significatio*, which "leaves more to be suspected than has actually been asserted" (Rhetorica ad Herennium 1954, pp. 400–401). Abbreviation deploys other clipping devices such as the ablative absolute, which has been likened to algebraic brackets in the way it sets apart an idea from the main sentence into an inner clause (Harris n.d.), or zeugma, where one verb does the work of many in a sentence, Puttenham's example being, "Fellowes and friends and kinne forsooke me quite" (Puttenham 1589, p. 136). The density of the abbreviated statement defamiliarizes its conventional appearance, bracketing off the expression from prosaic fullness, attracting the mind to make sense of the compressed words, and thereby requiring one to "see" the idea anew and recognize different aspects of its structure. In the same way in mathematics, an unexpected compression arrests the reader and requires an effort to achieve recognition.

Sometimes the pith of a simplification provides the mot juste. Sometimes it is periphrastic expansion of the expression that enables the reader to proceed to a solution of the problem. The eloquent mathematician knows which algebraic tropes to use and when in order to lead the reader to that moment when persuasion meets conviction, and opinion converges on knowledge. The efficacy of notation emerges in the ability of the mathematical sentence to expand and contract as the occasion requires, the virtuosity of the mathematician in gauging the right moment to substitute. In prudential rhetoric that feel for the occasion goes by various names: *kairos*, *to prepon*, *opportunitas*, decorum (Baumlin 2002). Often mischaracterized as a preoccupation with style rather than substance, rhetoric fundamentally concerns itself with fitness, with finding the right expression for the right moment.

3 Beautiful Notation

In the sixteenth, more explicitly in the seventeenth century, symbolic notation offered a new kind of visual experience that made one read and think differently. Just as perspective painting geometrized and thus rationalized the viewing subject's position and relationship with the art object, so symbolic notation made mathematicians reassess how graphically (that is, by writing rather than by diagram) to represent conceptual objects. In an apt analogy Frege noted that as his notation is to ordinary language so the microscope (invented late C16th) is to the eye (Frege 1970, p. 6). His claim applies also to early modern symbolic notation. When Thomas Harriot, in 1609, became the first person known to view a celestial object through a telescope, he understood for himself that this "Dutch trunke" not only enabled him to do better and faster what he already could do; it also generated new images and hence enabled him to ask questions not possible hitherto (Royal Astronomical Society 1996). Symbolic notation similarly amplified his horizon of possibility. One could not simply think *with* the "great invention" of algebra, to use Cardano's phrase, but *through* it.

If notation amplifies, it also abbreviates. In the Preface to the Reader of his *Key of the Mathematics*, William Oughtred praised notation for its compressive power (Oughtred 1647; see also Pycior 1997, pp. 43–45; Williams 2016, pp. 191–192).

> Which Treatise being not written in the usuall syntheticall manner, nor with verbous expressions, but in the inventive way of Analitice, and with symboles or notes of things instead of words, seemed unto many very hard; though indeed it was but their owne diffidence, being scared by the newnesse of the delivery; and not any difficulty in the thing it selfe. For this specious and symbolical manner, neither racketh the memory with multiplicity of words, nor chargeth the phantasie with comparing and laying things together; but plainly presenteth to the eye the whole course and processe of every operation and argumentation.

Like the density of gold, the compressive power of algebraic notation raises the value of conceptual labor and makes mathematical thought precious. Thus, Oughtred, when speaking of reading the Ancient mathematicians, described a quasi-alchemical process by which he transformed base language into symbolic notation, “uncasing the Propositions and Demonstrations out of their covert of words, designed them in notes and species appearing to the very eye” (Oughtred 1647; Pycior 1997, p. 46). For a culture trained to associate eloquence with copiousness, the marvel of notation lay in its power to contract thought. Where to admirers of notation, its density increased its preciousness, to its detractors, the compression rendered notation doubly inefficient because it required the mind to work twice as hard, as Thomas Hobbes observed at the end of his fifth lesson to the Savilian Professors (Hobbes 1656, p. 54):

> I shall also add, that Symboles, though they shorten the writing, yet they do not make the Reader understand it sooner than if it were written in words. For the conception of the Lines and Figures (without which a man learneth nothing) must proceed from words, either spoken or thought upon. So that there is a double labour of the mind, one to reduce your Symboles to words (which are also Symboles) another to attend to the Ideas which they signifie . . . . And thus having examined your panier of Mathematiques, and finding in it no knowledge neither of Quantity, nor of Measure, nor of Proportion, nor of Time, nor of Motion, nor of any thing, but only of certain Characters, as if a Hen had been scraping there . . .

Precisely because notation adds another degree of separation from the concept under discussion, Hobbes, in his third lesson, found it to be no more than an arid formalism, not beautiful at all (Hobbes 1656, p. 23).

> Symboles are poor unhandsome (though necessary) scaffolds of Demonstration; and ought no more to appear in publique, than the most deformed necessary business which you do in your Chambers.

Love it or hate it, notation indisputably possesses brevity as its chief characteristic. Its density must thus be offset with precision or what Frege calls “perspicuity” (Übersichtlichkeit) (Frege 1879, pp. vii, 9; Frege 1970, pp. 8, 17; Schlimm 2017, p. 29). Notation must be abbreviated enough to see the statement as a complete and meaningful expression yet amplified enough to avoid ambiguity and to make demonstrable the sequence of its logical steps.

Finding the optimal point between brevity and perspicuity is not a requirement exclusive to notation; it is fundamental to aesthetics, as Aristotle observed in his *Poetics* when considering the ideal duration of a tragedy. The right duration should be long enough for the full sequence of causation to emerge yet short enough to apprehend as a single, complex action, as an aesthetic totality. Thus, to take *Oedipus the King* as the *locus classicus* of this principle (Aristotle being a great admirer of Sophocles), Oedipus's tragedy carries the impact it does because the action of the play is, on the one hand, sustained enough for him to see how discrete facts of his past are in fact causally connected yet, on the other, compressed enough to comprehend the truth in one terrible moment of recognition (*anagnoresis*). Recognition is a cognitive act that depends on the play's optimal duration, on criteria that are essentially aesthetic. To use a scholastic distinction, where the play's brevity enables understanding of the action's quiddity or "whatness," its magnitude enables understanding of its haecceity or "thisness." Sufficient distance enables one to apprehend what this action *is*, sufficient proximity to apprehend what *this* action is.

Aristotle's word for what I have termed "duration" is *megethos* [magnitude, size, greatness] (Aristotle 1995, 1449b, l. 24). Of course, there is no quantitative formula that could determine *megethos* in this context because the duration of the play time is neither exact nor constant and requires the same kind of qualitative judgment needed in prudential rhetoric, which estimates the right word for the right occasion. Ancient Greek often used different words to distinguish between the qualitative and quantitative aspects of a thing. Thus, although *kairos* and *chronos* both mean "time," the former refers to the opportune moment while the latter refers to sequential, measurable time. Compare *bios* and *zoe*, both meaning life, where *bios* means purposive existence and *zoe* bare existence (Kerenyi 1962, pp. 12–13). Unlike these linguistic couplets, the single word *megethos* does double duty to refer to qualitative and quantitative size. In Aristotle's *Poetics* it refers to qualitative judgment; in his *Metaphysics*, *megethos* means measurable quantity, whether length (*mekos*) in one dimension, breadth (*planos*) in two, or depth (*bathos*) in three (Aristotle 1933, 1020a, ll. 10–13). Early English shares with Greek this fluid relation between measurement and ethos. Middle English *mesure* means "calculation," "just enough," and "moderation"; that is, it encompasses in one word the purely quantitative, the capacity for sufficiency to the occasion, and the ability to maintain balance and observe restraint (Oxford English Dictionary Online 2017, measure (n.); Middle English Dictionary 2001, mesure (n.)). To *mesure* a thing meant to capture its ethical nature just as surely as it meant to take its vital statistics.

In the classical and medieval lexicon of quantification, the measurable and the beautiful (or in Aristotle's case, the tragic) mutually constitute each other. The three conditions of beauty, claims Aquinas are *integritas* (totality, wholeness), *proportio* (internal harmony between the parts), and *claritas* (radiance, brightness, luminosity) (Aquinas 1964, Ia 39.8, ad 1). In the correspondences I have been tracking, rhetorical abbreviation and notational brevity enable apprehension of a thing in its totality, of its *integritas* and *quidditas*; rhetorical amplification and notational perspicuity enable apprehension of a thing's internal organization, of its *proportio* and *haecceitas*. As for *claritas*, Aquinas's third condition of beauty, it relies on an epistemology, even a theology, of light that derives from Neoplatonist thought. As light emits energy, it diffuses itself and is radiant (Tatarkiewicz 1970, p. 29). Like light, beauty is self-evident (note the verb *videre* in "evident"). It is not something one needs to be persuaded about in the same way that no one argues that two and two equal four; they just do and either you "see" it or you don't. The *quod erat demonstrandum* of a proof shows that it has reached a point when inferences are no longer needed and the truth of the proposition proclaims itself. In the same way radiant beauty is deictic. It shows itself as an objective reality, it "evidences" itself, and the only proper response is, "Yes, I see" (Allen 2016, p. 151). To its admirers, notation's density and ductility made it shine like bright gold, beautiful indeed.

Despite its kinship with rhetoric, algebraic notation developed in early modern England as a form of writing that eschewed, as Seth Ward put it, the "confusion or perturbatio[n] of the fancy made by words" (Pycior 1997, p. 115). Time and again, its virtue was trumpeted as the ability to communicate without the equivocations and amphibologies of ordinary language. The measure of such confidence in notation's promise is given by John Wilkins's attempt to bypass the constraints of all dialects through devising a concept-script in which symbols stood not for words but for "*things* and *notions*" (Wilkins 1668, p. 13). Yet, at the same time, thinkers of the day preoccupied themselves equally with rationalizing, methodizing, and "polishing" the English language into an idiom apt for scientific inquiry (Sprat 1667, p. 41). As interested in the logical systematicity of ordinary language as of mathematics, John Wallis, a talented cryptographer (Pycior 1997, p. 105), wrote a *Grammar of the English Language*. The capacity of English to name without dissembling and to stay close to concrete things made it the language of choice for empirical pursuit. By virtue of its plainness, English served as the matrix or mother tongue in which notation's precious density of thought could gestate. All the more reason, then, to pay attention to the words they used to describe this great invention of algebra.

4 The Words They Used

In this last section, I analyze the semantic histories of the English keywords used by the early modern algebraists to develop a specialized terminology for their mathematical pasigraphy: *specious*, *symbole*, and *note* or *notation*. My etymological analysis does not aim to recover some originary meaning and enthrone it as the word's proper signification. Nor does it offer a selection of definitions from which any can be chosen at will to change the meaning of a word. Rather it troubles the assumption that notation merely behaves like a well-behaved tool: "By relieving the brain of all unnecessary work, a good notation sets it free to concentrate on more advanced problems" (Whitehead 1948, p. 39). Etymology makes opaque the transparency of a linguistic sign that appears to adequate one word with one thing. It sounds out the multiple, sometimes disjunctive meanings that reverberate in an apparently univocal word.

Specious: Oughtred, quoted above, refers to the "specious and symbolicall manner" of notation. The word today usually has pejorative meaning, so it jars to encounter it used in a positive sense. By it, Oughtred means the habit of using letters rather than numbers: "literal" quantities as distinct from "numerical" quantities (Oxford English Dictionary Online 2017, literal (adj.), I.4). In this context, "literal" and "specious" both refer to algebraic notation. The distinctive occurrence of "specious" originates with François Viète, who uses it in its logical sense to distinguish between species and individual (Oxford English Dictionary Online 2017, species (n.) II.8.b). Early in his introduction to the *Analytic Art*, Viète describes "symbolic logistic as [a logistic] that can be expressed through species or forms" (*Speciosa [logistice] quae per species seu formas exhibitur*) (Viète 2006, p. 17). The logical meaning of *species* as "type" is clear from his context but the word had a broad associative range in Latin. More generally *speciosa* meant "beautiful" and *species* "appearance" or "outward form." It was within this non-logical semantic field that the words *specious* and *species* first existed in English. Oughtred took a word that in English basically meant "beautiful" and gave it technical application.

Deriving from Latin *species* (appearance), which itself derived from *specere* (to behold), the English word "specious," first attested in the early fifteenth century, meant "pleasing appearance" (Oxford English Dictionary Online 2017), specious (adj.) 1.a; *Middle English Dictionary* specious (adj.)). Within a century, it meant bright or showy appearance (as in a bird's plumage); within another, it had distinguished itself from the "real," in a contrast between appearance and reality (Oxford English Dictionary Online 2017, specious (adj.) 1.b and 2). Shortly thereafter, within the seventeenth century, it could imply fair-seemingness in the pejorative sense, hence fallaciousness (Oxford English Dictionary Online 2017, specious (adj.) 3, 4). The algebraic sense, obsolete today, constitutes the adjectival form of the logical term *species*, and it is arresting that the first attested use of the word to mean "insincere" or "fallacious" should occur in the *Leviathan* of philosopher Thomas Hobbes (Oxford English Dictionary Online 2017, specious (adj.) 3a, 3b; Hobbes 1651, pp. 73, 110). Hobbes' opinion of symbolic notation was famously low, and it must have seemed to him a most felicitous double entendre to extend the meaning of the word "specious" from "apparent" to "falsely plausible." In one of his many sideswipes at the mathematician John Wallis, who was a lover of symbolic notation, Hobbes wrote: "And for that [treatise] of your Conic Sections, it is so covered over with the scab of symbols, that I had not the patience to examine whether it be well or ill demonstrated" (Hobbes 1656, p. 316).

Yet to Wallis and other fellow-admirers, Viète's choice of *species* (meaning "type") to refer to general number offered an incisive way to combine "within the symbolic system two concepts hitherto considered as opposites, the arbitrary and the fixed, or the one and the multiple or even, maybe more significantly, the unspecified and the singular" (Serfati 2010, p. 109). They intuited that notation could hold together within its tensile economy ideas that, when unfolded into ordinary language, seemed to contradict each other. Such notation must indeed have seemed *speciosa* (beautiful) and pleasing to the eye.

Symbol: In the passage quoted, Oughtred uses the terms "specious" and "symbolic(al)" interchangeably. Yet the meaning of "symbol" is by no means self-evident, the word's semantic trajectory being fitful and indirect. Deriving from the Greek verb, *symballein*, "to throw together," the *symbolon* functioned like a tally stick except that coins were broken into half. From this act of matching fragments to make a connection, the early Christian theologian Cyprian applied the word to the baptismal creed, a symbol that comprised the matching half and thus the distinguishing mark of a Christian. One of the word's earliest meanings in English was consequently "a formal authoritative statement or summary of the religious belief … of a particular church or sect; a creed or confession of faith," so one could speak of "the Symbole of the Apostles," meaning the Apostles' Creed (Oxford English Dictionary Online 2017, symbol (n.1) 1a). From the association with pithiness or succinctness developed in the sixteenth century a more general meaning of symbol as a formula, maxim, motto, or synopsis. By 1656, one could speak of a symbol as "a short and intricate riddle or sentence" (Oxford English Dictionary Online 2017, symbol (n.1) 1b). For Oughtred, then, writing in the 1640s, a symbol bore connotations of hermeneutic compression, and offered an apt enough name for notation.

A second sense of the word, independent but connected, emerged in the late sixteenth century (Oxford English Dictionary Online 2017, symbol (n.1) 2a):

> Something that stands for, represents, or denotes something else (not by exact resemblance, but by vague suggestion, or by some accidental or conventional relation); esp. a material object representing or taken to represent something immaterial or abstract, as a being, idea, quality, or condition; a representative or typical figure, sign, or token.

Key here is the idea of substitution of one entity for another. The first attested occurrence of "symbol" as an English word with substitutionary meaning comes in poetry, in Edmund Spenser's *Faerie Queene*. The knight Sir Guyon, who represents temperance in Spenser's elaborate allegory, encounters a dying woman, who has stabbed herself from grief at being betrayed, and who daubs her infant son in her blood as a "pledge" to her unblemished chastity (Spenser 2013, II.i.37). Guyon vows to avenge her death and tries to wash the infant in the nearby fountain but the stains remain. The Palmer (or pilgrim) who accompanies Guyon explains why. The waters are pure and will not permit admixture with the spilt blood, so the infant must remain with bloodied hands as a "Sacred Symbole" of his mother's innocence and reminder to avenge her (Spenser 2013, II.ii.10). Resonant with allusion to Pilate's hand-washing declaration of innocence, this enigmatic passage is difficult to interpret. Rejected by the pure water, the blood implies impurity, yet the Palmer explicitly interprets the blood as a mark of innocence, hence purity (as well as a goad to vengeance). As a symbol, the bloody hands become overdetermined, their multiple meanings—impurity, innocence, and a reminder of vengeance—potentially contradicting each other. "The symbol has condensed, collapsing within itself disjunctive and identificatory understandings of the sign, forcing the reader's attention onto its 'mystery'" (Tribble 1996, p. 32).

At this point algebraic symbolism, which deals in precision, seems to part company with poetic symbolism, which deals in polysemy. Yet Viète's "specious" notation shares with poetry the power to hold together seeming disjunctives in a "dialectic," as Serfati describes it. Values that are both arbitrary and fixed in notation "appear as contradictions in natural language" (Serfati 2010, p. 110). The generality of an algebraic expression enabled an equation's values to be true at all times, unlike the rule of false position, which muddled its way to a solution by means of determinate values that were technically incorrect. In algebra, “the variable is always correct until and through the moment when it can be replaced by a determinate number” (Williams 2016, p. 200). Just so, the poetic symbol holds together meanings that, in the contingent points of narrative, denote variously. In this way symbolic expression denotes differently from pedestrian language. Ideas that contradict in daily vernacular become coherent and fruitful within the symbol’s referential universe.

Notation: The word “note” seemed to promise some freedom from complicity with poetics and aesthetics. Compact and cryptic, a note had the advantage of being distinct from letters, which mimicked sounds and thus individual languages, as John Wilkins observed: “Besides this common way of Writing by the ordinary *Letters*, the Ancients have sometimes used to communicate by other *Notes*, which were either for *Secrecy*, or *Brevity*” (Wilkins 1668, p. 12). (Wilkins, by the way, preferred the term “character” for the concept-script he developed.) Leibniz’s Latin term for what we call symbol was *nota* (Bochenski 1961, pp. 274–275). Present as early as the period of Old English (Oxford English Dictionary Online 2017, note (n.2) I.1.a), “note” from the outset suggested a distinguishing mark other than an alphabetic one.

From “note” derives “notation.” Its mathematical meaning, namely “the process or method of representing numbers, quantities, relations, etc., by a set or system of signs or symbols, for the purpose of record or analysis,” only dates from the early eighteenth century, and applies subsequently to music, dance, chemistry, logic, chess, linguistics, and so on (Oxford English Dictionary Online 2017, notation (n.) 6). “Notation” as a mathematical term of art constitutes a second-order nominalization from the concrete to the abstract, as if to claim the term exclusively for a specialized concept-script, but the word has an older history in English, its earliest occurrence, from the mid-sixteenth century, coming in a treatise of grammar and rhetoric (Sherry 1555, pp. xxv–xxvi):

> By notacion, that is, when by certain markes, and signes we do describe any thing: as, if a man understandyng anger, wil saye it is the boylyng of the minde, which bringeth palenes unto the countenaunce, burning to the iyes, tre[m]bling to the partes of the body. Also, when for the proper name we put the country, the sect, or some great act: as. For Virgil, the poet of Mantua. For Aristotle, the prince of peripateticall schole. For Scipio, the destroyer of Carthage and Numance.

The word began its life in English as a rhetorical trope, a name for the kind of circumlocution or “speaking around” that captured in description the nature of a thing or person. In the rhetorical treatises, Latin *notatio* went hand in glove with *effictio*, together constituting a complete *descriptio* of persons. Where *effictio* dwelt on the corporeal features of the person described, *notatio* sketched the quintessential character of a person, and was a kind of *ethopoeia* (character description). It delineated a person in pithy, well-chosen details, omitting anything extraneous. All of these rhetorical terms, *notatio* included, classify as kinds of *enargeia* that selectively represent the object of description. The more vividly they do their job, the more they unravel the distinction between description and naming. Notation as a rhetorical term already came associated in the minds of these classically trained, early modern mathematicians with a descriptive act that distilled the very essence of the object of reference into the mark of signification. It is ironic then that when John Wallis refers to the "Notes or *Symbols*" of algebra saving one from a "multitude of Words, and long Periphrases" (Pycior 1997, p. 113), he had already mentally sundered algebraic notation from rhetorical *notatio* and forgotten their shared genealogy.

5 Conclusion

Like Theophrastus's character sketches, notation functions as a kind of outline or diagram of the concept. Where Hobbes saw only random hen scrapings, these early modern mathematicians saw marks that could illustrate rather than merely denote a concept. When Recorde introduces his equals sign to save himself tedium he describes his parallels "as gemowe [twin, from Latin *gemini*] lines of one length, thus = because no 2 things can be more equal" (Recorde 2010, p. 242). Thomas Harriot appears to have built his notation for inequality on the foundation of Recorde's parallels (Tanner 1962, pp. 166–167). Although not all mathematical *notes* were motivated in this way, their creators held these symbols to standards of decorum and beauty just as poets and rhetoricians did their words. Compact where words tended to be copious, early modern mathematical notation observed its own aesthetic criteria, paying attention to the spatial management and visual logic of the page. In doing so, symbolic notation positioned mathematical objects and operations in a textual practice as richly polysemous as poetry, even if its squiggles only looked, to Hobbes and his ilk, like hen scrapings.

References

Allen, Valerie (2016) "Chaucer and the Poetics of Gold", in Beatrice Fannon (ed.) *New Casebooks: Reading Medieval Literature*, London: Palgrave, pp. 144—159.

Aquinas, Thomas, St. (1964-1975) *Summa Theologiae*, ed. and trans. Thomas Gilby, Oxford: Blackfriars.

Aristotle (1933, repr. 1989) *Metaphysics, Books I—IX*, ed. and trans. Hugh Tredennick, Cambridge, MA: Harvard University Press.

Aristotle (1995) *Poetics*, ed. and trans. Stephen Halliwell, Cambridge, MA: Harvard University Press.

Baltzell, Jane (April 1967) "Rhetorical 'Amplification' and 'Abbreviation' and the Structure of Medieval Narrative" *Pacific Coast Philology* 2, pp. 32—39.

Baumlin, James S. (2002) "Ciceronian Decorum and the Temporalities of Renaissance Rhetoric", in Phillip Sipiora and James S. Baumlin (eds) *Rhetoric and Kairos: Essays in History, Theory, and Praxis*, Albany: State University of New York Press, pp. 138-164.

Bochenski, I.M. (1961) *A History of Formal Logic*, rev. and trans. Ivo Thomas, Indiana: Notre Dame University Press.

Cajori, Florian (1928—1929), *A History of Mathematical Notations*. 2 vols. Chicago: Open Court, repr. in one vol. (2012) Mineola, NY: Dover.

Carroll, Lewis (2013) *Alice in Wonderland*, ed. Donald J. Gray, 3rd edn, New York and London: Norton.

Cifoletti, Giovanna (2006) "Mathematics and Rhetoric. Introduction", *Early Science and Medicine* 11.4, 369—389.

Faral, Edmond (1924, repr. 1962) *Les arts poétiques du XIIe et du XIIIe siècle: recherches et documents sur la technique littéraire du moyen âge*, Paris: Champion.

Frege, Gottlob (1879) *Begriffsschrift eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Louis Nebert.

Frege, Gottlob (1970) *Begriffsschrift* (English), in *Frege and Gödel: Two Fundamental Texts in Mathematical Logic*, ed. Jean van Heijenoort, Cambridge MA: Harvard University Press.

Harris, William. http://www.orbilat.com/Languages/Latin/Alternative_Grammars/Harris_Grammar/Latin-Harris_17.html

Heeffer, Albrecht (2009) "On the Nature and Origin of Algebraic Symbolism", in Bart van Kerkhove (ed.), *New Perspectives on Mathematical Practices: Essays in Philosophy and History of Mathematics*, Singapore, World Scientific Publishing, pp. 1—21.

Heeffer, Albrecht and Maarten Van Dyck, eds, (2010) *Philosophical Aspects of Symbolic Reasoning in Early Modern Mathematics*, Studies in Logic 26, London: College Publications, King's College.

Heeffer, Albrecht (2015) "On the Historiography of Mathematical Notations", in Douglas R. Furman and Albrecht Heeffer (eds), *Writing the History of Mathematical Notation 1483-1700* by Sister Mark Leontius Schulte, Boston, MA: Docent Press, pp. 3—43.

Hobbes, Thomas (1651) *Leviathan, or The Matter, Form, and Power of a Common-Wealth Ecclesiastical and Civil*, London: printed for Andrew Crooke, at the Green Dragon in St. Paul's Church-yard.

Hobbes, Thomas (1656) *Six Lessons to the Professors of the Mathematiques*, London: printed by J.M. for Andrew Crook, at the Green-Dragon in Paul's Church-yard.

Kerenyi, C. (1962) *The Religion of the Greeks and Romans*, trans. Christopher Holme, New York: Dutton.

Lanham, Richard A. (1991) *A Handlist of Rhetorical Terms*, 2nd edn, Berkeley: University of California Press.

Lloyd, G.E.R. (1990) *Demystifying Mentalities*, Cambridge: Cambridge University Press.

Middle English Dictionary (2001), University of Michigan http://quod.lib.umich.edu/m/med/

Moxon, Joseph (1700) *Mathematicks made easie. or, A mathematical dictionary explaining the terms of art and difficult phrases used in arithmetick, geometry, astronomy, astrology, and other mathematical sciences*, 3rd edn, London: printed for J. Moxon at the *Atlas* in *Warwick-Lane* and Thomas Tuttell at the *King's Arms* and *Globe* at *Charing-Cross*.

Oughtred, William (1647) *The Key of the Mathematicks New Forged and Filed: Together with a Treatise of the Resolution of all kinde of Affected Æquations in Numbers*, London: printed by Tho. Harper for Rich. Whitaker.

Oxford English Dictionary Online (2017), Oxford: Oxford University Press.

Puttenham, George (1589) *The Arte of English Poesie*, London: printed by Richard Field, dwelling in the black-Friers, neere Ludgate.

Pycior, Helena M. (1997) *Symbols, Impossible Numbers, and Geometric Entanglements: British Algebra through the Commentaries on Newton's Universal Arithmetick*, Cambridge: Cambridge University Press.

Recorde, Robert (2010) *The Whetstone of Witte, A Facsimile*, ed. Thomas Gordon Roberts, Mickleover, Derby: Renascent Books.

Rhetorica ad Herennium (1954) ed. and trans. Harry Caplan, Cambridge, MA: Harvard University Press.

Rotman, Brian (2000) *Mathematics as Sign: Writing, Imagining, Counting*, Stanford: Stanford University Press.

Royal Astronomical Society (1996-2015) https://www.ras.org.uk/news-and-press/news-archive/68-news2009/1641-celebrating-thomas-harriot-the-worlds-first-telescopic-astronomer-ras-pn-0947

Schlimm, Dirk (Published online. 2017) "On Frege's *Begriffsschrift* Notation for Propositional Logic: Design Principles and Trade-offs", *History and Philosophy of Logic*, 1—34.

Serfati, Michel (2010) "Symbolic revolution, scientific revolution: mathematical and philosophical aspects", in Heeffer and Van Dyck, eds, *Philosophical Aspects*, pp. 103—22.

Sprat, Thomas (1667) *The History of the Royal-Society of London for the Improving of Natural Knowledge*, London: printed by T.R. for J. Martyn and J. Allestry.

Spenser, Edmund (2013) *The Faerie Queene*, rev. 2nd edn, ed. A.C. Hamilton, Hiroshi Yamashita, and Toshiyuki Suzuki, London and New York: Routledge.

Sherry, Richard (1555) *A Treatise of the figures of Grammer and Rhetorike*, profitable for al that be studious of eloquence, and in especiall for suche as in grammer scholes doe reade moste eloquente poetes and oratours. London: in aedibus Ricardi Totelli.

Tanner, R.C.H. (1962 Dec.) "On the Role of Equality and Inequality in the History of Mathematics" *British Journal for the History of Science* 1.2, pp. 159—69.

Tatarkiewicz, Wladyslaw (1970) *History of Aesthetics, Vol. II: Medieval Aesthetics*, The Hague and Paris: Mouton.

Tribble, Evelyn B. (1996) "The Partial Sign: Spenser and the Sixteenth-Century Crisis of Semiotics," in Douglas F. Rutledge (ed.), *Ceremony and Text in the Renaissance* (Cranbury, NJ: Associated University Presses), pp. 23—34.

Viète, François (2006) *The Analytic Art*, trans. T. Richard Witmer, Mineola, NY: Dover.

Whitehead, Alfred North (1948, repr. 1972) *An Introduction to Mathematics*, Oxford: Oxford University Press.

Wilkins, John (1668) *An Essay towards a Real Character, and a Philosophical Language*, London: printed for Sa. Gellibrand and John Martyn.

Williams, Travis D. (2016) "Mathematical *Enargeia*: The Rhetoric of Early Modern Mathematical Notation", *Rhetorica* 34(2), 163—211.

Polar Ordinates in Bernoulli and L'Hôpital

Robert E. Bradley

Abstract Priority for the invention of polar coordinates should probably be accorded to Newton, although Jakob Bernoulli has priority of publication in 1691, because Newton's results were only published posthumously, almost half a century later. However, it was not until the middle of the 18th century that polar coordinates took on the form that is used today. Earlier versions all featured ordinates emanating from a single point or pole, with some geometric object playing the role that now belongs to an angular coordinate. The largest and most accessible collection of these early schemes of polar ordinates is probably to be found in the Marquis de L'Hôpital's *Analyse des infiniment petits* (1696), based on the lessons given to the Marquis by Johann Bernoulli. We describe Bernoulli's approaches to polar ordinates, as presented in l'Hôpital's textbook.

Keywords Polar coordinates • Pole • Ordinate • Analytic geometry • Differential calculus • Newton • Bernoulli • L'Hôpital • Euler • Spiral of Archimedes • Conchoid of Nichomedes • Parabolic Spiral

1 Introduction

Polar coordinates first appeared in their modern form in *Introductio in analysin infinitorum* Euler (1748) by Leonhard Euler (1707–1783). Prior to that, various systems that are essentially equivalent were used by mathematicians on a more or less *ad hoc* basis, in order to solve problems related to particular curves. All of these schemes involved a *polar ordinate* – a line emanating from a fixed point, or *pole*, to a generic point on the curve in question – and one other geometric quantity, which played a role roughly equivalent to that of the angular coordinate of the modern system.

R.E. Bradley (✉)
Department of Mathematics and Computer Science, Adelphi University,
Garden City, NY 11530, USA
e-mail: bradley@adelphi.edu

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*,
Proceedings of the Canadian Society for History and Philosophy of Mathematics
La Société Canadienne d'Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_5

Isaac Newton (1642–1727) seems to be the first person to have used a coordinate system essentially equivalent to polar coordinates; see (Boyer 1949). This innovation was first published in *The Method of Fluxions* (Newton 1736), the posthumous English translation of his treatise *Artis analyticae specimina vel Geometria analytica* (Whiteside 1969, p. 28–253), written in or around 1671. We will examine Newton's coordinate system in section 2. Even though Newton's discovery was published in 1736, it remained "hidden in plain sight" until Boyer's article (1949) in the middle of the 20th century. Prior to this, Jakob Bernoulli (1654–1705) was generally granted priority for the discovery of polar coordinates, on the basis of two articles published in *Acta Eruditorum* (Bernoulli 1691, 1694). It's evidently impossible that Bernoulli would have been seen Newton's unpublished treatise and, conversely, Boyer argues that Newton's use of the polar system was not a later interpolation into the manuscript of his work. Consequently, this would appear to be a genuine case of independent discovery.

In the same way that the independent invention of the calculus by Newton and Gottfried Leibniz (1646–1716) rested upon the work of many earlier mathematicians of the 17th century – arguably going all the way back to antiquity – Newton and Bernoulli's polar systems were to some extent foreshadowed in the works of other mathematicians earlier in the same century. Coolidge (1952) examines some of these precursors in an article that followed Boyer's "discovery" of Newton's polar system.

Newton's work went unpublished until 1736 and the readership of *Acta Eruditorum* was limited. Moreover, authors in *Acta* often wrote in a style that hid, rather than exposed, the ideas behind their discoveries. Leibniz' early articles on the differential and integral calculus (Leibniz 1684, 1686) were notoriously obscure and Bernard de Fontenelle (1657–1757) wrote of the calculus that even in 1696,

> ...the new mathematics was nothing but a kind of mystery and, so to speak, a cabalistic science shared among five or six people. They often gave their solutions in the journals, without revealing the method that produced them. Even when one could discover the method, it was nothing but a few feeble rays of this science that had escaped, and the clouds quickly closed again. (Fontenelle 1708, p. 133–134; translation in Bradley et al. 2015, p. 299)

The Newton/Bernoulli's polar system became accessible to a broader mathematical public through the efforts of the Marquis de l'Hôpital (1661–1704), who wrote the textbook *Analyse des infiniment petits* (L'Hôpital 1696). There is a direct line from Jakob Bernoulli to de l'Hôpital, running through Johann Bernoulli (1667–1748), Jakob's younger brother. Jakob Bernoulli collaborated closely with his younger brother during the early 1690's and it was Johann who taught the calculus to the Marquis, providing him with a framework for his textbook, as well as many of the results that were contained therein.

Polar ordinates, in some form or another, are used in studying four different curves in de l'Hôpital's *Analyse*. Additionally, there is some effort to give a general treatment of them that is independent of the particulars of each individual coordinate system. We will examine l'Hôpital's use of these ordinates in section 4 of this paper. References to the *Analyse* will be to the English translation (Bradley et al. 2015).

2 Newton's Coordinate System for Spirals

Newton composed a treatise on quadrature and infinite series (*De Analysi per æquationes numero terminorum infinitas*) and put it into private circulation in the summer of 1669. Although he was encouraged to publish it, he chose first to make major revisions, including the addition of a complete exposition of his method of fluxions. He was appointed Lucasian Professor of Mathematics in October 1669, and so the revised treatise (*Artis analyticae specimina vel Geometria analytica*) wasn't completed until 1671. Following a series of delays and disputes, Newton ultimately decided not to publish the work; see (Whiteside 1969, p. 1–12). Nine years after Newton's death, John Colson (1680–1760) published an English translation of the work (Newton 1736), to which he appended some 200 pages of commentary.

In his *Method of Fluxions*, Newton first explains how to expand algebraic equations into infinite series. He then describes the method of finding fluxions and inverse fluxions. Applications follow, the finding of maxima and minima, and of tangents to curves. In the section on finding tangents, he first uses rectangular coordinates, but then describes eight additional methods of finding tangents. Essentially, he is describing alternate coordinate systems, beginning with oblique coordinates.

Newton described his "Seventh Manner: For Spirals" in a single page (§49–53). He used Figure 1 to illustrate his method. The given spiral, whose tangents we seek, is ADE. The point A is the center of a circle BG. Newton imagined the radius AB to be revolved around the center A, while the point D is moving along the radius,

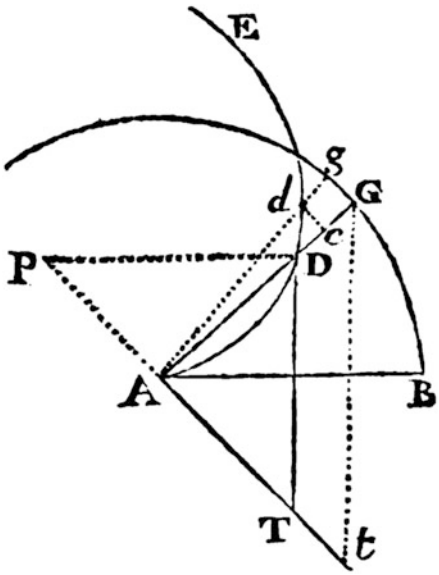


Fig. 1 Newton's polar coordinates in *Method of Fluxions*: $x = BG$ and $y = AD$

thereby tracing out the spiral. In Newton's words, D may "be conceived to move any how, so as to describe the spiral ADE." He let $x = BG$ and $y = AD$, and supposed that the spiral is given by an equation relating x and y.

In an infinitely small increment of time, the radius moves from AG to Ag and the tracing point moves from D to d. We draw c on AG so that $Ac = Ad$. The fluxions $\dot{x}$ and $\dot{y}$ are then the infinitely small increments Gg and Dc, respectively. Let AT be drawn perpendicular to the ordinate AD and suppose that the tangent DT has already been drawn. Then by similar triangles, $cd : cD :: AT : AD$. By similar sectors, $cd : Gg :: Ad : AG :: AD : AG$, in the second case because the difference between Ad and AD is infinitely small. Now suppose that Gt is drawn parallel to DT, so that by similar triangles, $AD : AG :: AT : At$, so that $cd : Gg :: AT : At$. Combining this with $cd : cD :: AT : AD$, we have $Gg : cD :: AD : At$, or $\dot{y} : \dot{x} :: AD : At$.

To construct the tangent then, we calculate the fluxions $\dot{x}$ and $\dot{y}$ from the given equation relating x and y and locate t on AT such that $\dot{y} : \dot{x} :: AD : At$. Finally, the line DT, drawn parallel to Gt, is the tangent to the curve ADE at D. The method is illustrated with three examples, including the Spiral of Archimedes, in which $\frac{ax}{b} = y$ for two given constants a and b.[1]

In this passage, Newton has more or less given us polar coordinates. His polar ordinate y is identical with our radial coordinate r. His abscissa x is $c\theta$, where θ is the usual angular coordinate and c is the radius of the given circle. Although this scheme identifies every point in the plane and thus could be used to describe any curve, Newton apparently only conceived of using it for spirals. In the case of the Spiral of Archimedes, the equation $y = \frac{ax}{b}$ takes the modern polar form $r = k\theta$, where $k = \frac{ac}{b}$, under the substitutions $y = r$ and $x = c\theta$.

3 The Bernoulli Brothers and the Marquis de L'Hôpital

Jakob Bernoulli was clearly experimenting in various ways with radial coordinates during the early 1690s. His publications using these schemes are the reason many authors gave him priority for the invention of polar coordinates; see (Boyer 1949, fn. 1).

In 1691, his paper on the parabolic spiral (Bernoulli 1691) appeared in *Acta Eruditorum*. Bernoulli used radial ordinates in this treatment of the spiral but, unlike polar coordinates, his ordinates were erected as normals on a given circle, extending inwards to the point where they intersected the spiral. Like Newton, he used abscissas x that were arcs on a given reference circle, such as the arcs BC and BD in Figure 2 or Figure 3, but his ordinate y extended inward, such as CF and DG.

[1]A modern text might simply write $y = ax$, but Newton gave the equation in a homogenous form, which arises from the proportional relationship $y : x :: a : b$.

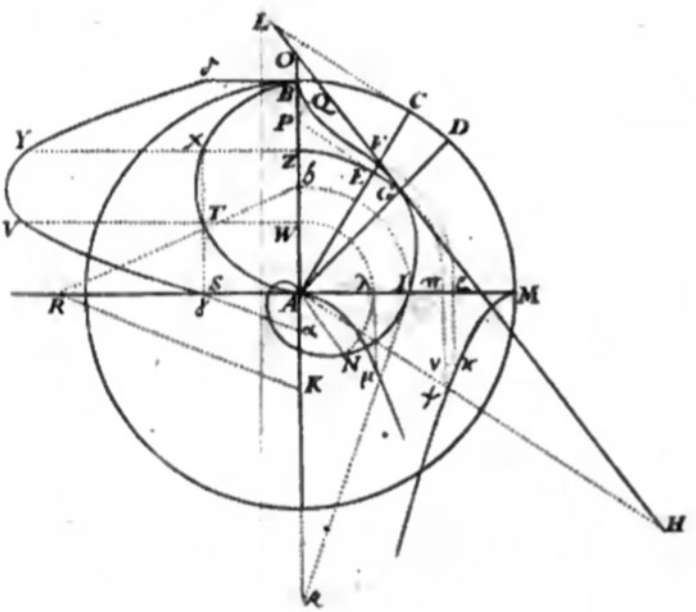


Fig. 2 Bernoulli's coordinates in 1691: $x = BC$ and $y = CF$

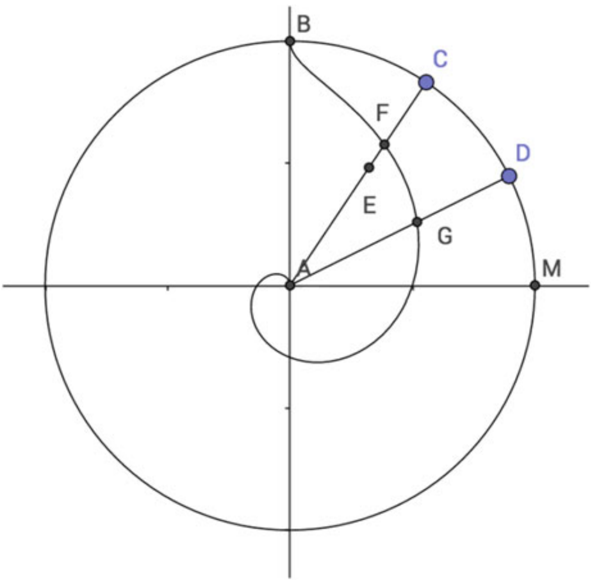


Fig. 3 A simplified version of Bernoulli's 1691 figure, showing the parabolic spiral, the reference circle, and the ordinates

Thus he had $x = c\theta$ and $y = c - r$, where r and θ are the modern polar coordinates and c is the radius of the reference circle. Boyer (1949), Coolidge (1952), and others consider this to be the first published appearance of polar coordinates.

Three years later, Bernoulli published a paper in *Acta Eruditorum* concerning the curvature of elastic laminæ (Bernoulli 1694), in which he used the same coordinates that Newton had employed when calculating the radius of curvature: y for the polar ordinate r and x for the arc $c\theta$, where c is the radius of the osculating circle.

When Leibniz first published his papers on the differential and the integral calculus (Leibniz 1684, 1686), they were not widely understood. Together, Jakob and Johann Bernoulli studied these early papers in the late 1680s and extended the scope of the calculus in the 1690s, although they cooperated less and less with each passing year, until their increasingly quarrelsome relationship disintegrated completely in about 1697. Meanwhile, Johann met the Marquis de l'Hôpital in Paris in November 1691 and tutored him the new calculus over the course of almost a year, first in the Marquis' Paris apartment and later as a guest at his *Chateau* in Oucques; see (Bradley et al, p. vii–xiv). Bernoulli's lessons, as well as further material he provided the Marquis through their subsequent correspondence, formed the basis of *Analyse des infiniment petits*, the first differential calculus textbook.

The Bernoulli brothers were obviously making use of the polar ordinates of (Bernoulli 1694) well in advance of 1694, because Johann used the same coordinate system in the lessons he gave the Marquis de l'Hôpital in his Paris apartment during the winter of 1691–92. Fortunately for the historical record, l'Hôpital insisted that Bernoulli commit those lessons to paper. Bernoulli typically composed each lesson, which he wrote in Latin, the night before he gave them to the Marquis. Luckily, Bernoulli was sharing rooms with his friend and later colleague at the University of Basel, Johann Heinrich von Stähelin (1668–1721). Von Stähelin made copies of the lessons before Bernoulli handed them over to the Marquis.

Bernoulli also kept copies of these lessons. L'Hôpital had incorporated the first part, on the differential calculus, into the first four chapters of the *Analyse*, but Bernoulli published the larger second part, concerning the integral calculus, in his collected works (Bernoulli 1742, p. 385–558). In 1921, Paul Schafheitlin discovered a manuscript copy of the full set of lessons, both the differential and integral, in the library of the University of Basel. Schafheitlin published the first portion (1922) and argued in his introduction that the manuscript was a copy made in 1705 by Bernoulli's nephew Nikolaus (1687–1759), who had been living with him in Groningen. Because the latter part of this manuscript is a near-perfect match with what Bernoulli had published in 1742, we can be quite certain that the first part is essentially the same set of lessons Bernoulli gave to l'Hôpital in 1691–92. An English translation of Bernoulli's lessons on the differential calculus is available in (Bradley et al. 2015, p. 187–231).

Problem XI in Bernoulli's lessons to l'Hôpital is "To Find the Tangent in the Spiral of Archimedes" (Bradley et al, p. 204–205). To solve this problem, Bernoulli used the same polar system that Newton used in *The Method of Fluxions* and his brother had employed in (1694). Like Jakob, he oriented his arcs in a clockwise direction from a fixed point on the circumference of the reference circle, whereas

Newton had gone counterclockwise. Johann used letter x to denote the polar ordinate, unlike his brother and Newton, who had labeled it y, and he did not use any letter to denote the abscissa or arc on the reference circle.

4 Polar Ordinates in L'Hôpital's *Analyse*

The Marquis de l'Hôpital published his *Analyse des infiniment petits* in 1696. The first four chapters followed the lessons he had been given by Johann Bernoulli in broad outline, but provided more general propositions and additional examples. A comparison of the texts reveals a successful collaboration between a mathematical creator (Bernoulli) and a talented expositor. Although l'Hôpital acknowledged "having received much from the illuminations of Messrs. Bernoulli, particularly those of the younger, presently Professor at Groningen," he never specifically admitted just how much his textbook owed to Johann Bernoulli, both for its superstructure and many of its particular results. Although Bernoulli claimed authorship of the book after the Marquis' death in 1704, he made no protest at the time of publication, because l'Hôpital had paid him a stipend in return for his silence and the permission to publish some of his results, including one of the cases of what we now call L'Hôpital's Rule (Bradley et al. 2015, p. xiii-xvi).

In his first chapter, l'Hôpital gave a very lucid exposition of the rules of the differential calculus, in the manner that Bernoulli had taught him. He followed this with a very long chapter concerned with finding tangents to curves, which serves as a sort of catalog of many of the curves that were known to mathematicians in the late 17th century. For three of these curves, l'Hôpital used some form of polar ordinates.

4.1 *The Spiral of Archimedes*

L'Hôpital treated the Spiral of Archimedes as a particular example of a more general construction given in Proposition V of Chapter 2 (Bradley et al. 2015, p. 20–21), where a curve *FMD* is defined by means of a pole *F* and a reference curve *APB*; see Figure 4. A generic point *M* on the curve *FMD* is given by drawing the line *FP* from the pole to the reference curve and marking the point *M*, where "the relation of the part FM to the portion of the curve *AP* is given by whatever equation we wish." L'Hôpital used the letter y to denote the polar ordinate *FM* and x to denote the abscissa *AP*.

In the case of the Spiral of Archimedes, l'Hôpital lets the reference curve *APB* be a circle of radius a with its center at F. If b is the arc length of some portion of the circle *APB*, the equation $y = \dfrac{ax}{b}$ defines the Spiral of Archimedes. L'Hôpital's coordinates x and y are precisely the same as those used by Jakob

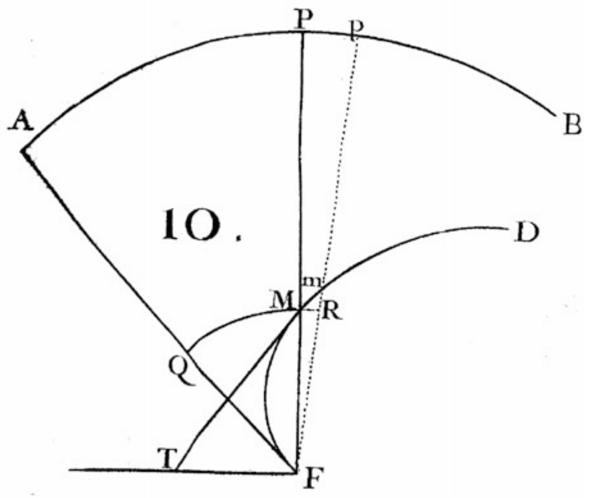


Fig. 4 L'Hôpital's Figure 10, the Spiral of Archimedes: $x = AP$ and $y = FM$

Bernoulli (1694) and Newton (1736). He differed from Newton in that his arcs were oriented clockwise from the initial point A whereas Newton went counterclockwise from his initial point (point B in Figure 1). There is another difference between l'Hôpital and Newton's treatment. Both gave equations that are equivalent to the modern polar equation $r = k\theta$ but, whereas Newton's scheme allows k to be defined independently of the radius of the reference circle, in l'Hôpital's treatment, the spiral will always intersect the reference circle at a point who distance from the initial point A is $b = \frac{a}{k}$. In other words, l'Hôpital's scheme gives a constant of proportionality of $k = \frac{a}{b}$.

4.2 The Conchoid of Nicomedes

Analogously to the Spiral of Archimedes, l'Hôpital found tangents to the Conchoid of Nicomedes as a special case of a more general Proposition VI of Chapter 2 (Bradley et al. 2015, p. 22–23), for any curve which is given in a particular way in terms of a polar ordinate. We will illustrate his construction with a figure taken from later in the *Analyse*, where he found the inflection points of the Conchoid of Nicomedes.

The Conchoid of Nicomedes is the curve AFK in Figure 5. It is defined in terms of a pole P and a straight line BC, which is called the *directrix* or *asymptote* of the curve. Figure 5 shows only the right half of the curve; the left side of the curve is symmetric about the line ABP. A generic point F on the curve is defined by drawing

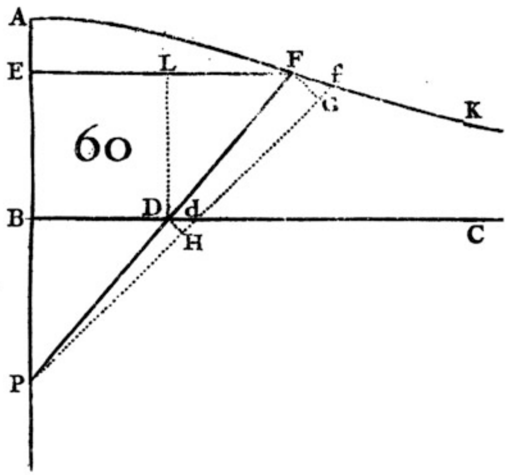


Fig. 5 L'Hôpital's Figure 60, the Conchoid of Nicomedes: $x = PD$ and $y = PF$

the line PD from the pole to the directrix, which is then extended beyond D to the point F, such that $DF{=}BA$. The Conchoid of Nicomedes has two parameters, $a = BA$ and $b = PB$.

In l'Hôpital's general treatment in Proposition VI, the straight line BC may be any curve for which we know how to draw tangents. He denoted the polar ordinate PF by y and the distance PD by x, and allowed the relationship between these two to be "expressed by any equation." Because in the case of the Conchoid of Nicomedes the portion DF has a constant length a, the equation relating y and x in this case is simply $y - x = a$.

The modern polar equation for the curve AFK in Figure 5 can be found by taking the origin to be the pole P and the y-axis to be the line PA, so that the directrix is the line $y = b$. The radial angle θ is then equal to $\angle PDB$, so that $\csc\theta = \frac{x}{b}$. Therefore l'Hôpital's coordinates are $x = b\csc\theta$ and $y = r$, which gives the modern polar equation for the Conchoid of Nicomedes: $r = a + b\csc\theta$.

We should note that Newton also described such a system of coordinates in *The Method of Fluxions*, which he called his "Fourth Manner" for finding tangents (Newton 1736, p. 55).

4.3 The Logarithmic Spiral

The Logarithmic Spiral is the curve with polar equation $r = ae^{b\theta}$. Unlike the two previous examples, this curve was not known in classical times. The Logarithmic Spiral has many interesting properties, including the *equiangular* property: at any point on the curve, the line joining that point to the center always makes the same angle with the tangent at that point.

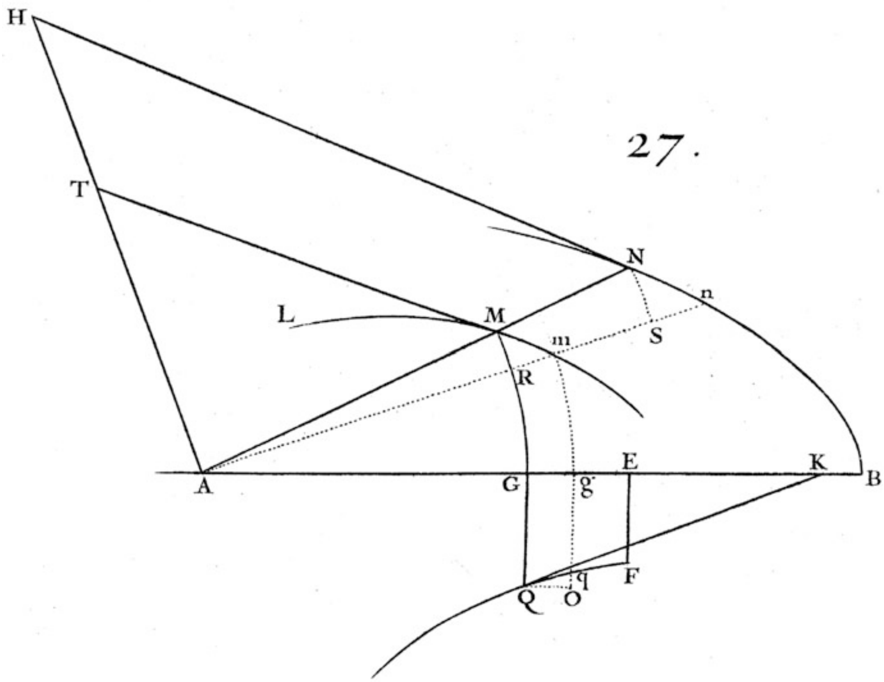


Fig. 6 L'Hôpital's Figure 27, the Logarithmic Spiral: $y = AM$

Similarly to the previous examples, l'Hôpital proved a general Proposition XIII in Chapter 2 (Bradley et al. 2015, p. 38–40), which applied to the Logarithmic Spiral as a special case. The general case is illustrated in Figure 6. The curve *BN* is a reference curve which, in the case of the Logarithmic Spiral is a circle with center *A*. The curve *FQ* is a second reference curve, which in the case of the Logarithmic Spiral is a hyperbola. A generic point *M* on the curve is marked on the radius *AN* in such a way that the area of the circular sector *ANB* has a particular relation to the area *EGQF* under the second reference curve. Given the appropriate point *G* on the radius *AB*, the point *M* on *AN* is such that $AM = AG$. Because l'Hôpital did not have direct access to the transcendental functions e^x or $\ln x$, he instead used the area under the hyperbola to give him indirect access to logarithms. His condition on the area *EGQF* was that it be half of the area of the sector *ANB*. In this case, he was able to show that the resulting spiral had the equiangular property and was therefore a logarithmic spiral.

As in the previous examples, l'Hôpital used the variable y to denote the polar ordinate *AM*, which is equal to the r of modern polar coordinates. In this case, it is the area of the circular sector *ANB*, which of course is proportional to the arc *BN*, which plays the role of the θ.

4.4 Concavity and Polar Ordinates

We've seen that in Chapter 2 of the *Analyse*, l'Hôpital introduced three different coordinate systems which, despite their differences, all had the common feature of a polar ordinate emanating from a fixed point.

Chapter 2 was followed by a chapter on maximum/minimum problems. Then in Chapter 4, l'Hôpital introduced higher differentials and used second order differentials to determine the inflection points and cusps of a curve. In his investigation, he identified two cases. The first case was Cartesian coordinates, in which he showed that at an inflection point or cusp, the second differential d^2y (which he denoted ddy) must take the value zero or infinity. Modern readers may imagine formally dividing d^2y by $(dx)^2$ and will then recognize this as the familiar condition of the second derivative changing sign at an inflection point, or of being undefined at a cusp.

L'Hôpital's second case was that in which the curve "has as its ordinates the straight lines BM …, which all emanate from the same point B" (Bradley et al. 2015, p. 69). That is, any of the three polar schemes which were used in Chapter 2 were included in this second case. This situation is illustrated in Figure 7. L'Hôpital also had a Figure 57, illustrating the same case in a curve convex towards the axis.

In these situations, he let $y = BM$ be the polar ordinate, as he had done in all the examples of polar ordinates in Chapter 2, and $dx = MR$, the perpendicular from the point M on the curve to a polar ordinate Bm that is infinitely close to the ordinate BM. Because this straight line is infinitely small it may, according to the rules of the differential calculus, be considered to be an infinitesimal arc of the circle of radius y with center at B. Making use of modern polar coordinates, we have $y = r$ as always, and $dx = r\,d\theta$. L'Hôpital didn't ever give an analytic expression for x, neither here nor anywhere else in the *Analyse*. Instead, he solved for dx in the differential triangle MRm using similar triangles, and thereby was able to find expressions involving dx, dy, and higher differentials.

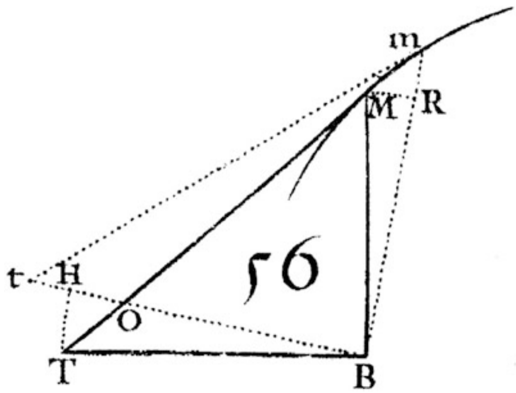


Fig. 7 L'Hôpital's Figure 56, a curve concave towards the axis: $y = BM$ and $dx = MR$

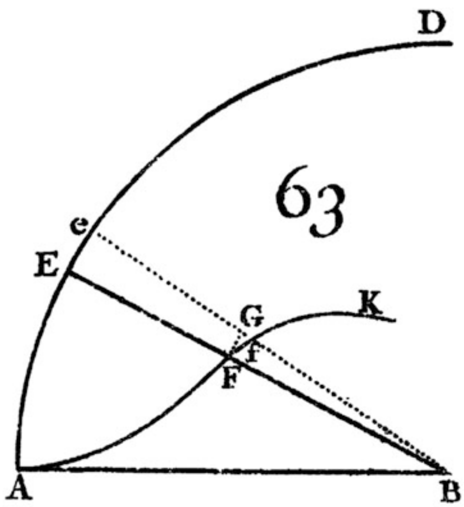
Having set up this coordinate system, l'Hôpital then showed that where a curve has an inflection point or cusp, we must have $dx^2 + dy^2 - y\,ddy$ equal either to zero or to infinity.[2] He went on to use the same coordinate system in Chapter 5, where he calculated the radius of curvature.

4.5 The Parabolic Spiral

In modern polar form, the equation of the Parabolic Spiral is $(r-a)^2 = c^2\theta$, which has two branches $r = a \pm c\sqrt{\theta}$. The Spiral of Fermat is usually understood to be the special case $a = 0$. The curve has one inflection point, which is at the origin when $a = 0$, but otherwise is removed from the origin. In §73 of Chapter 4 of the *Analyse* (Bradley et al. 2015, p. 76), l'Hôpital considered the problem of finding this inflection point.

L'Hôpital began with the 17th century geometric definition of the parabolic spiral. The circle *AED* in Figure 8 has radius a and center at B. A generic point F on the spiral is marked on the radius BE so that the square on FE is equal to the product of the arc length of AE and a given constant b. This is why it was natural for Jakob Bernoulli to consider a coordinate system where AE is the abscissa and EF is the ordinate for this curve in (1691). However, l'Hôpital used the polar ordinate BF in this problem, consistently with previous examples. He let $BF = y$ and the

Fig. 8 L'Hôpital's Figure 63, the Parabolic Spiral: $z = AE$ and $y = BF$



[2]The case $dx^2 + dy^2 - y\,ddy = 0$ may be translated into a modern polar form using the substitutions $y = r$ and $dx = r\,d\theta$, then formally dividing through by $(d\theta)^2$. The resulting condition is $\frac{d^2r}{d\theta^2} = r + \frac{1}{r}\left(\frac{dr}{d\theta}\right)^2$.

arc $AE = z$,[3] so that his equation for the spiral was $(a - y)^2 = bz$. Because $y = r$ and $z = a\theta$ in modern terms, his spiral satisfies the polar equation $(r - a)^2 = c^2\theta$, where $c = \sqrt{ab}$, the geometric mean of two parameters.

In Figure 8, dx and dz are the arcs FG and Ee, respectively. By similar sectors, we have $\frac{dx}{dz} = \frac{y}{a}$. Taking differentials of the equation $(a - y)^2 = bz$, we have $2(a - y)(-dy) = b\,dz$, or $b\,dz = 2y^2\,dy - 2ay\,dy$. Combining these, we have

$$dx = \frac{y\,dz}{a} = \frac{2y^2\,dy - 2ay\,dy}{ab}.$$

Substituting this into the formula $dx^2 + dy^2 - y\,ddy = 0$, l'Hôpital found a quintic equation in the variable y. It is interesting to note that the value of y corresponding to the inflection point couldn't be given explicitly, but was implicitly given as the solution of this equation.

5 Concluding Remarks

We've seen four different coordinate systems given by l'Hôpital in his *Analyse*. Of course, they are originally due to the Bernoulli brothers, primarily transmitted to him by Johann, as l'Hôpital in fact acknowledged in his preface. All the schemes involved a fixed pole from which a polar ordinate y emanated. The other coordinate varied from scheme to scheme, representing some geometric quantity, whether the length of an arc, the length of a segment, or the area of a circular sector. With hindsight and the benefit of modern polar coordinates, we can recognize each of them as a simple function of the angular coordinate θ.

The modern polar coordinate system, with variables r and θ related to the Cartesian coordinate by the equations $x = r\cos\theta$ and $y = r\sin\theta$, is due to Euler and first appeared about half a century after l'Hôpital's *Analyse* in (Euler 1748, vol. 2, Chapter XVII), although in this first occurrence, Euler used z for the radial coordinate and ϕ for the angular coordinate. That Euler should have done this is hardly surprising, given that it was he who first worked out trigonometric functions in their modern form, as functions of arcs on a unit circle. It's a short step from the coordinates of Newton and the Bernoullis, using arcs $c\theta$ of a circle of arbitrary radius, to adopting the uniform standard of the unit circle. The benefit of doing so would presumably be very clear to the person who turned trigonometric lines into functions of the unit circle.

[3]His choice of z instead of x, which he had used for the same arc in the Spiral of Archimedes, is explained by his polar coordinate scheme as described in the previous section.

References

Bernoulli Jakob (1691) Specimen calculi differentialis in dimensione parabolæ helicoidis. Acta eruditorum 10:13–22

Bernoulli Jakob (1694) Curvatura laminæ elasticæ. Acta eruditorum 13:262–276

Bernoulli Johann (1742) Opera Omnia, vol. 3. Bousquet, Lausanne

Boyer C (1949) Newton as an Originator of Polar Coordinates. Amer Math Monthly 56:73–78

Bradley R, Petrilli S, Sandifer C (2015) L'Hôpital's Analyse des infiniment petits. Birkhäuser, Cham

Coolidge J (1952) The Origin of Polar Coordinates. Amer Math Monthly 59:78–85

Euler L (1748) Introductio in analysin infinitorum. Bousquet, Lausanne

Fontenelle B de (1708) Histoire du renouvellement de l'Académie royale des sciences. Boudot, Paris

Leibniz G (1684) Nova methodus pro maximis et minimis. Acta eruditorum 3:467–373

Leibniz G (1686) Geometria recondita et analysi indivisibilium Acta eruditorum 5:292–300

L'Hôpital G de (1696) Analyse des infiniment petits. Imprimerie Royale, Paris

Schafheitlin Paul (1922) Johannis (I) Bernoullii Lectiones de calculo differentialium. Verhandlungen der Naturforschenden Gesellschaft in Basel 34:1–32

Newton I (trans. Coulson J) (1736) The Method of Fluxions and Infinite Series. Woodfall, London

Whiteside D, ed. (1969) The Mathematical Papers of Isaac Newton, vol. 3. Cambridge University Press, Cambridge

Magic Squares of Subtraction of Adam Adamandy Kochański

Henryk Fukś

Abstract The problem of the construction of magic squares occupied many mathematicians of the 17th century. The Polish Jesuit and polymath Adam Adamandy Kochański studied this subject too, and in 1686 he published a paper in *Acta Eruditorum* titled "Considerationes quaedam circa Quadrata et Cubos Magicos". In that paper he proposed a novel type of magic square, where in every row, column, and diagonal, if the entries are sorted in decreasing order, the difference between the sum of entries with odd indices and those with even indices is constant. He called them *quadrata subtractionis*, meaning squares of subtraction. He gave examples of such squares of orders 4 and 5, and challenged readers to produce an example of square of order 6. We discuss the likely method which he used to produce squares of order 5, and show that it can be generalized to arbitrary odd orders. We also show how to construct doubly even squares. At the end, we show an example of a square of order 6, sought by Kochański, and discuss the enumeration of squares of subtraction.

Keywords Magic squares of subtraction • Enumeration of magic squares • 17th century mathematics • Jesuit mathematicians • Polish mathematicians

1 Introduction

Contributions of Jesuits to science and mathematics have attracted considerable attention in recent years (Feingold 2013; O'Malley et al 2016). It is now well documented that the Society of Jesus played an especially important role in the intellectual life of Polish-Lithuanian Commonwealth (Stasiewicz-Jasiukowa 2004), producing a number of notable writers, preachers, theologians, missionaries, and scientists. The Polish Jesuit Adam Adamandy Kochański (1631–1700) was one of

H. Fukś (✉)
Department of Mathematics and Statistics, Brock University, St. Catharines, ON, Canada
e-mail: hfuks@brocku.ca

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*,
Proceedings of the Canadian Society for History and Philosophy of Mathematics
La Société Canadienne d'Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_6

such remarkable individuals. He pursued problems in a multitude of diverse fields, including philosophy, mathematics, astronomy, philology, design and construction of mechanical clocks and computing devices, and many others. He published relatively little, and his published works remain rather unknown today, even among historians of mathematics and science. In recent years, however, his accomplishments seem to have attracted some more interest, in large part due to the efforts of B. Lisiak SJ, who reprinted all the published works of Kochański (2003a,b,c) and who authored a comprehensive account of Kochański's life and work (Lisiak 2005). Moreover, B. Lisiak SJ and L. Grzebień SJ gathered and published all the surviving letters of Kochański, including correspondence with Gottfried Leibniz, Athanasius Kircher SJ, Johannes Hevelius, Gottfried Kirch, and many other luminaries of the 17th century (Lisiak and Grzebień 2005).

The most important and original mathematical works of Kochański appeared in *Acta Eruditorum* between 1682 and 1696. Among those, three are particularly interesting, namely papers published in *Acta* in 1682, 1685, and 1686 (all three reprinted in Kochański 2003a, English translation in Fukś 2015). The 1685 paper, *Observationes cyclometricae*, is relatively well known and is often considered to be the most interesting one. It proposes for the rectification of the circle an approximate procedure which received some attention from both contemporaries of Kochański and historians of mathematics (Montucla 1754; Cantor 1880; Günther et al 1921). What is generally less known, however, is that the same paper included an interesting sequence of rational approximations of π, somewhat similar to continued fractions. Detailed discussion of this sequence and its properties can be found in a recent article (Fukś 2012), so we will not consider it here. We will only mention that the sequence of integers on which the aforementioned approximation of π is based is now included in the Online Encyclopedia of Integer Sequences as A191642 (OEIS Foundation Inc. 2015). I proposed to call it *Kochański's sequence*.

The subject of this note is the 1686 paper, *Considerationes quaedam circa Quadrata & Cubos Magicos*. This work, like all of Kochański's other papers, touches upon more than one subject. In the first part, "classical" magic squares are discussed, and Kochański presents some previously unknown magic squares using the method of their construction developed by A. Kircher. The second part is more original and interesting, because Kochański introduces there a new type of magic square, to be called *quadrata subtractionis* or *squares of subtraction*. An $n \times n$ square of subtraction is an arrangement of consecutive integers from 1 to n^2 in such a way that in every row, column, and diagonal, if the entries are sorted in decreasing order, the difference between the sum of entries with odd indices and the sum of entries with even indices is constant. Kochański gives examples of $n = 4$ and $n = 5$ squares of subtraction, and then challenges fellow mathematicians to produce an example of a square of order 6. To my knowledge, nobody has ever taken this challenge, and nobody has ever studied magic squares of subtraction in the 330 years following the publication of *Considerationes*. The only author mentioning magic squares of subtraction in a published work is Z. Pawlikowska (1969), who briefly discussed them in a paper describing the mathematical works of Kochański and who guessed the most likely method used by him to produce squares of order 5.

The purpose of this article, therefore, is to bring magic squares of subtraction back to life, with the hope of stimulating others to study their properties and methods of construction. We will take a closer look at the examples supplied by Kochański and present the general method of construction of squares of odd order and doubly even order, based on these examples. We will also give an example of an $n = 6$ square, which he requested, and discuss the enumeration of squares of subtraction.

2 Notation and Examples

Kochański defines $n \times n$ squares of subtraction as arrangements of the consecutive integers $1, 2, \ldots, n^2$ which have the same *residuum* in rows, columns, and diagonals. His definition of “residuum” is as follows.

> *In his Quadratis Subtractione tractandis ita proceditur. Numerus in assumpta Columna, Trabe, vel Diagono minimus, subducendus est a proxime majore, residuum majoris hujus detrahatur a proxime consequente, atque ita porro : ultimum enim residuum est illud universale [...].*
>
> In these Squares produced by Subtraction one proceeds as follows. The smallest number in a selected Column, Row, or Diagonal is subtracted from the next-largest number, the difference is subtracted form the subsequent larger number, and so on: the final result is the universal number [...].

This means, for example, that if numbers in a given row, column, or diagonal of a square 5×5, sorted in decreasing order, are a_1, a_2, a_3, a_4, a_5, then the aforementioned residuum is

$$a_1 - (a_2 - (a_3 - (a_4 - a_5))).$$

Obviously, this can be written in a more convenient way, without nested parentheses. Let us, therefore, define the residuum of a vector $\mathbf{x} = (x_1, x_2, \ldots x_n)$ using modern notation. First, we sort the vector $\mathbf{x}$ in decreasing order, obtaining $\tilde{\mathbf{x}}$ such that $\tilde{x}_1 \geq \tilde{x}_2 \geq \ldots \geq \tilde{x}_n$; then the *residuum* of $\mathbf{x}$ is defined as

$$\mathrm{res}(\mathbf{x}) = \tilde{x}_1 - \tilde{x}_2 + \tilde{x}_3 - \ldots \tilde{x}_n = \sum_{i=1}^{n} (-1)^{i+1} \tilde{x}_i. \tag{1}$$

A Magic square of subtraction is an arrangement of numbers in a square array such that all rows, columns, and both diagonals have the same residuum. If entries of the array belong to the set $\{1, 2, \ldots, n^2\}$, then the magic square of subtraction will be called *normal*. In what follows we will consider only normal squares (unless indicated otherwise), so we will omit the designation “normal” and we will refer to them simply as *squares of subtraction*. The number n will be called the *order* of the square of subtraction. Figure 1 shows an example of a square of subtraction of order 4 with residuum equal to 8.

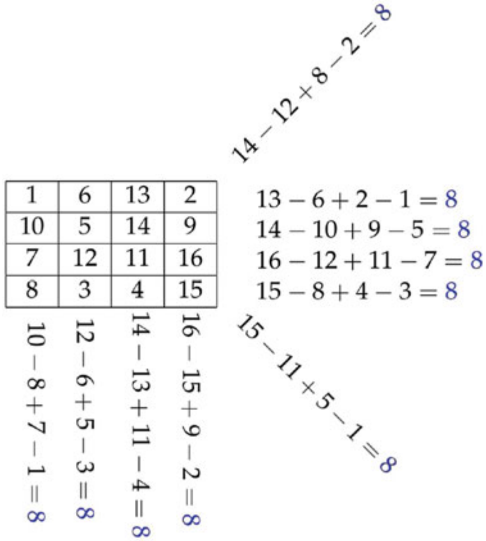


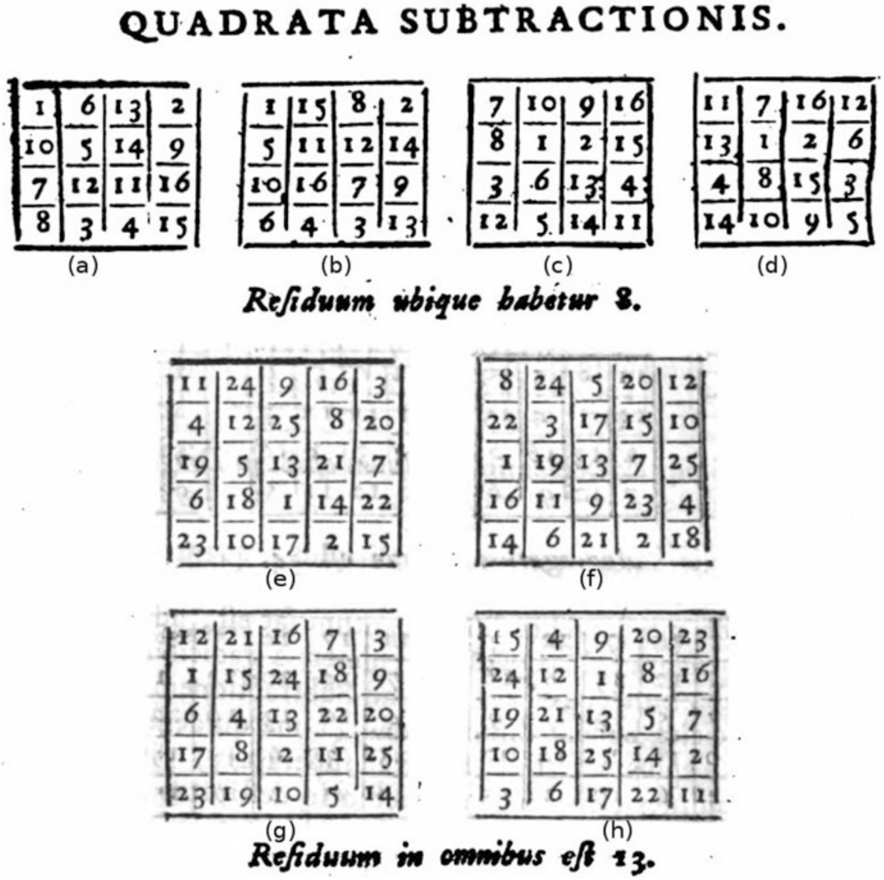
Fig. 1 Example of a square of subtraction of order 4

This is actually one of the squares given by Kochański. In the original paper he gave examples of seven other squares of subtraction, of orders 4 and 5, as shown in Figure 2. It is worth noticing that the smallest order he considered is 4, and that he does not mention any smaller orders. He was likely aware that squares of order 2 and 3 do not exist, and this is not too hard to prove.

3 Formal Construction for Odd *n*

We will start with odd-order squares, as these are easier to construct. Kochański gives examples of squares of order 5, and this is indeed the smallest possible odd order. The first of these, shown in Figure 2e, is reproduced again in Figure 3, as the framed square. As observed by Pawlikowska (1969), it can be constructed by writing consecutive numbers from 1 to 25 along diagonal lines of length 5, one below the other, with alternating direction. Afterwards, one needs to relocate the numbers which are outside the framed square to its interior, by "wrapping" them back to the square (boldface numbers in Figure 3 are those which have been relocated).

How did Kochański discover this method of construction? From his surviving letters we know that he was familiar with works of Athanasius Kircher SJ, whom he greatly esteemed and with whom he maintained correspondence (Lisiak and Grzebień 2005; Lisiak 2005). In Kircher's book *Arithmologia* (Kircher 1665), a method of construction of magic squares of summation is discussed, here reproduced in Figure 4. In this method, we write consecutive integers along diagonal lines, one below the other, but maintaining the same direction. Numbers which fall outside of the square $n \times n$ are relocated to the interior of the square. This method



QUADRATA SUBTRACTIONIS.

(a)

1	6	13	2
10	5	14	9
7	12	11	16
8	3	4	15

(b)

1	15	8	2
5	11	12	14
10	16	7	9
6	4	3	13

(c)

7	10	9	16
8	1	2	15
3	6	13	4
12	5	14	11

(d)

11	7	16	12
13	1	2	6
4	8	15	3
14	10	9	5

Reſiduum ubique habetur 8.

(e)

11	24	9	16	3
4	12	25	8	20
19	5	13	21	7
6	18	1	14	22
23	10	17	2	15

(f)

8	24	5	20	12
22	3	17	15	10
1	19	13	7	25
16	11	9	23	4
14	6	21	2	18

(g)

12	21	16	7	3
1	15	24	18	9
6	4	13	22	20
17	8	2	11	25
23	19	10	5	14

(h)

15	4	9	20	23
24	12	1	8	16
19	21	13	5	7
10	18	25	14	2
3	6	17	22	11

Reſiduum in omnibus eſt 13.

Fig. 2 Original examples of squares of orders 4 and 5 from Kochański's *Considerationes*

Fig. 3 Construction of a square of subtraction of order 5. Relocated entries are shown in boldface

				1				
			10		2			
		11	**24**	9	**16**	3		
	20	**4**	12	**25**	8	**20**	4	
21		19	**5**	13	**21**	7		5
	22	**6**	18	**1**	14	**22**	6	
		23	**10**	17	**2**	15		
			24		16			
				25				

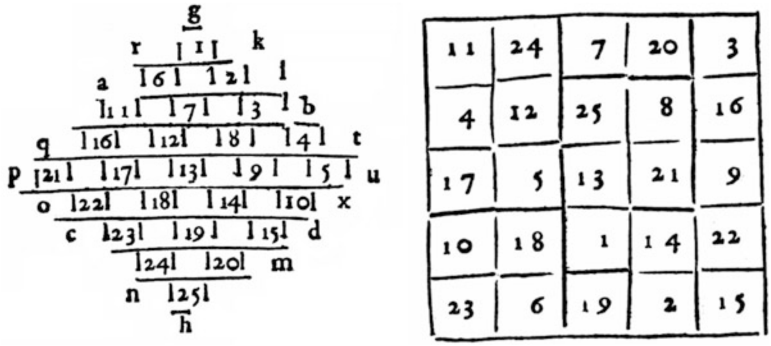


Fig. 4 Construction of a summation magic square of order 5 reproduced from pp. 89 and 90 of Kircher's *Arithmologia* (Kircher 1665)

has been originally proposed by Claude Gaspard Bachet de Méziriac (1624), but since Kochański does not mention Bachet's name in *Considerationes*, it seems likely that he learned about it from Kircher (Kircher's name is explicitly listed in the first paragraph of *Considerationes* as one of the "ingenious men" who have studied magic squares).

Kochański noticed that by reversing the order in which we write numbers on each diagonal in Kircher's example, instead of a magic square of summation we obtain a square of subtraction. He did not prove this fact, but merely produced examples. We will now show that this method indeed produces the desired magic squares of subtraction, and that it can be generalized to any odd order n. To simplify notation, we will define

$$[k]_n = \Big((k-1) \mod n\Big) + 1. \tag{2}$$

Note that if $k \in \mathbb{Z}$ then $[k]_n \in \{1, 2, \ldots, n\}$. This will be used to bring indices of matrices back to the range $\{1, 2, \ldots, n\}$.

Proposition 1 *Let n be an odd integer, $n \geq 3$, $m = (n+1)/2$, and let*

$$f(i,j) = (m-i)(-1)^j + n(n-j) + m. \tag{3}$$

Then the matrix

$$M_{i,j} = \begin{cases} f\left(\frac{1}{2}(i+j), \frac{1}{2}(j-i) + m\right) & \text{if } i,j \text{ are of the same parity,} \\ f\left(\left[\frac{1}{2}(i+j+n)\right]_n, \left[\frac{1}{2}(j-i+n) + m\right]_n\right) & \text{otherwise.} \end{cases} \tag{4}$$

is a magic square of subtraction of order n with residuum $\frac{n^2+1}{2}$.

We will be using Figure 3 to illustrate the proof. Let us first explain where the formulae in equations (3) and (4) come from. Note that the magic square of order 5 shown in Figure 3 has been obtained by rotating the matrix

$$A = \begin{pmatrix} 21 & 20 & 11 & 10 & 1 \\ 22 & 19 & 12 & 9 & 2 \\ 23 & 18 & 13 & 8 & 3 \\ 24 & 17 & 14 & 7 & 4 \\ 25 & 16 & 15 & 6 & 5 \end{pmatrix} \tag{5}$$

by 45 degrees counterclockwise. The function $f(i,j)$ defined in the proposition is simply the indexing function used to construct A, so that $A_{i,j} = f(i,j)$. After A is rotated, one needs to "squeeze" all entries of this matrix into a square box 5×5, shown in Figure 3. "Squeezing" is achieved by relocating all entries which are outside the box to the inside of the box, by bringing their indices to the range $\{1, 2, \ldots, 5\}$ with $[\cdot]$ operator. This simply means that each entry located outside the box at (i,j) is relocated to $([i]_n, [j]_n)$. The result is the matrix M,

$$M = \begin{pmatrix} 11 & 24 & 9 & 16 & 3 \\ 4 & 12 & 25 & 8 & 20 \\ 19 & 5 & 13 & 21 & 7 \\ 6 & 18 & 1 & 14 & 22 \\ 23 & 10 & 17 & 2 & 15 \end{pmatrix}, \tag{6}$$

which is the desired magic square of subtraction. The rotation by 45 degrees around the origin can be described algebraically as the transformation $(i,j) \to \left(\frac{1}{2}(i+j), \frac{1}{2}(j-i)\right)$, and we need to add m to the second index to bring the origin of coordinate system to the right location, thus

$$M_{i,j} = f\left(\frac{1}{2}(i+j), \frac{1}{2}(j-i) + m\right). \tag{7}$$

Note that the above works only for entries which ended up inside the box, not needing relocation. These happen to be entries for which i, j are both odd or both even. The remaining entries must be relocated, and this is achieved by combined translation and use of the operator $[\cdot]_n$, resulting in the second line of equation (4).

From Figure 3, one can see that the third row of M is the diagonal of A. Other rows of M can be constructed from numbers lying on two diagonal segments of A, as illustrated in Figure 5.

This can be easily generalized to arbitrary n. For convenience, we will introduce index $p \in \{0, 1, \ldots n-1\}$ labeling rows of M, such that the actual row number is equal $[p-m-1]_n$. The middle row of M will then always correspond to $p = 0$. The set

$$R_p = \{f(p+i, i)\}_{i=1}^{n-p} \cup \{f(p+i-n, i)\}_{i=n-p+1}^{n}, \tag{8}$$

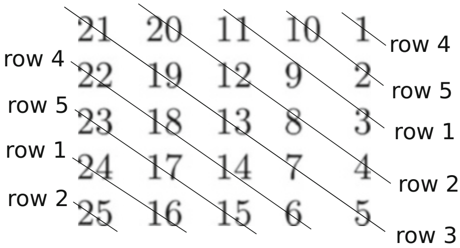


Fig. 5 Correspondence between diagonals of A and rows of M

composed of two diagonal lines of A, contains all entries of the row $[p-m-1]_n$ of M. Note that the second component of the above set is empty when $p=0$, as in this case we only need one diagonal line (the main diagonal).

It is important to notice that the elements of the set R_p are already sorted in descending order, thus we can easily compute its residuum,

$$\operatorname{res}(R_p) = \sum_{i=1}^{n-p}(-1)^{i+1}f(p+i,i) + \sum_{i=n-p+1}^{n}(-1)^{i+1}f(p+i-n,i) \tag{9}$$

Using the definition of f (equation 3) and the fact that n is odd, one can easily compute the above sums, obtaining

$$\operatorname{res}(R_p) = \frac{n^2+1}{2}. \tag{10}$$

This means that the residua of all rows of M are the same and equal to $\frac{n^2+1}{2}$.

Exactly the same observations can be made for columns of M. We label them with the integer $q \in \{0,1,\ldots,n-1\}$, so that the actual column number is $[3-q]_n$. Then all the entries of the column $[3-q]_n$ of M are contained in the set

$$C_q = \{f(n-i-q+1,i)\}_{i=1}^{n-q} \cup \{f(2n-i+q+1,i)\}_{i=n-q+1}^{n}. \tag{11}$$

As before, the above set is already sorted in descending order, thus

$$\operatorname{res}(C_q) = \sum_{i=1}^{n-q}(-1)^{i+1}f(n-i-q+1,i) + \sum_{i=n-q+1}^{n}(-1)^{i+1}f(2n-i+q+1,i). \tag{12}$$

Computing the sums we obtain

$$\operatorname{res}(C_q) = \frac{n^2+1}{2}, \tag{13}$$

confirming that the residuum is the same for every column of M.

What is now left is checking the diagonals of M. The main diagonal of M corresponds to the middle column of A, or m-th column, $\operatorname{diag} M = \{f(i,m)\}_{i=0}^{n}$. It is sorted in increasing order, thus

$$\operatorname{res}(\operatorname{diag} M) = \sum_{i=0}^{n} (-1)^i f(i,m). \tag{14}$$

The antidiagonal of M corresponds to the m-th column of A, $\operatorname{adiag} M = \{f(m,i)\}_{i=0}^{n}$ and it is sorted in decreasing order, thus

$$\operatorname{res}(\operatorname{adiag} M) = \sum_{i=0}^{n} (-1)^{i+1} f(m,i). \tag{15}$$

Both sums in equations (14) and (15) can be easily computed. They are both equal to $(n^2+1)/2$, thus M is indeed a magic square of subtraction. □

Proposition 1 produces, for $n = 5$, the square shown in Figure 2e. How did Kochański obtain the remaining three squares of order 5, shown in Figures 2f, g, and h? He does not explain it in the paper, but it is rather easy to guess, following observations made by Pawlikowska (1969).

Keeping the notation used in the proof of Proposition 1, let us recall that the magic square M is obtained by rotating the initial matrix A and then relocating outside entries to the interior of the square. Obviously, if A is transposed before rotation and relocation, the resulting square will still be magic. Transposition of A, therefore, does not affect the magic property. One can also show that any permutation of columns of A which keeps the middle column in the same place, performed before the rotation and relocation, preserves the magic property of the resulting square M (recall that the middle column of A becomes the diagonal of M, so it cannot be moved). The same applies to any permutation of rows of A keeping the middle row in the same place.

The remaining three 5th order squares of Figure 2 can be obtained if certain operations of the aforementioned type are applied to rows and columns of matrix A of equation (5) before it is rotated. These are shown below.

- For the square shown in Figure 2f: transpose A, then apply the permutation $(1,2,3,4,5) \to (4,5,3,1,2)$ to the rows of A^T and the permutation $(1,2,3,4,5) \to (4,1,3,5,2)$ to its columns;
- For the square shown in Figure 2g: apply the permutation $(1,2,3,4,5) \to (2,5,3,1,4)$ to the rows of A and the permutation $(1,2,3,4,5) \to (1,4,3,2,5)$ to its columns;
- For the square shown in Figure 2h: apply the same permutation $(1,2,3,4,5) \to (5,2,3,4,1)$ to both the rows and columns of A.

$$\begin{bmatrix} 28&47&16&37&12&35&4\\ 5&27&48&17&36&11&29\\ 30&6&26&49&18&42&10\\ 9&31&7&25&43&19&41\\ 40&8&32&1&24&44&20\\ 21&39&14&33&2&23&45\\ 46&15&38&13&34&3&22 \end{bmatrix} \begin{bmatrix} 37&78&35&66&21&62&15&46&5\\ 6&38&79&34&65&22&63&14&54\\ 53&7&39&80&33&64&23&55&13\\ 12&52&8&40&81&32&72&24&56\\ 57&11&51&9&41&73&31&71&25\\ 26&58&10&50&1&42&74&30&70\\ 69&27&59&18&49&2&43&75&29\\ 28&68&19&60&17&48&3&44&76\\ 77&36&67&20&61&16&47&4&45 \end{bmatrix}$$

Fig. 6 Squares of subtraction of order 7 and 9

The above operations will produce three matrices,

$$A' = \begin{bmatrix} 7&10&8&6&9\\ 4&1&3&5&2\\ 14&11&13&15&12\\ 24&21&23&25&22\\ 17&20&18&16&19 \end{bmatrix},\ A'' = \begin{bmatrix} 22&9&12&19&2\\ 25&6&15&16&5\\ 23&8&13&18&3\\ 21&10&11&20&1\\ 24&7&14&17&4 \end{bmatrix},\ A''' = \begin{bmatrix} 5&16&15&6&25\\ 2&19&12&9&22\\ 3&18&13&8&23\\ 4&17&14&7&24\\ 1&20&11&10&21 \end{bmatrix}.$$

By rotating each one of them and relocating outside entries to the interior we obtain the magic squares of subtraction shown, respectively, in Figure 2f, g, and h.

Obviously, since Proposition 1 is valid for any odd n, we can use it to construct squares of subtraction of higher orders. Figure 6 shows squares of subtraction of order 7 and 9 obtained by this method.

4 Even Order

Just as in the case of magic squares of summation, magic squares of subtraction of even order are more difficult to construct. Kochański gave four examples of squares of order 4, the smallest order for which magic squares of subtraction exist, and these are shown in Figure 2a–d. He did not explain how they were constructed. Close inspection of them reveals certain regularities in the arrangement of numbers, and from this one can guess how the construction probably proceeded.

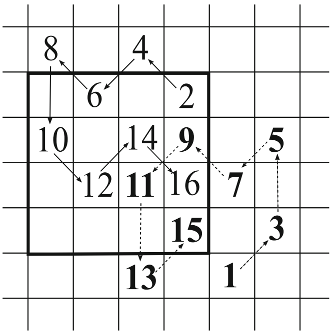
Most likely, his first observation was that the arrangement of odd and even numbers must follow some pattern reflecting constrains required for the square to be magic. The simplest of such patterns is shown below.

odd	even	odd	even
even	odd	even	odd
odd	even	odd	even
even	odd	even	odd

It satisfies the obvious requirement that the residua of all rows, columns, and diagonals must have the same parity, and that the total number of odd entries must be the same as the total number of even entries. He then probably proceeded, by trial and error, to fill this pattern with consecutive odd and even integers, and one of the most natural ways to do this would be to follow a "zigzag" path. Such a path does not need to start inside the square, just as in the case of squares of order 5. After some tinkering with such paths, one discovers that the placement of odd and even numbers shown in Figure 7 fits the bill. This requires, as in the case of order 5 square, the relocation of numbers which are outside the square to its interior, using "periodic boundary conditions", that is, left-right and bottom-top wrapping. One thus discovers the square shown in Figure 2a.

The square of Figure 2b can be produced by applying to the first square the transformation shown schematically in Figure 8. By direct verification one can check that this indeed produces a magic square of subtraction.

Finally, if we apply to both rows and columns of the squares shown in Figure 2a and 2b the permutation $(1, 2, 3, 4) \rightarrow (3, 1, 4, 2)$, we will obtain the remaining two



$\longrightarrow$

1	6	13	2
10	5	14	9
7	12	11	16
8	3	4	15

Fig. 7 Construction of the first 4-th order square of subtraction of Figure 2. Numbers outside of the framed square are relocated to the interior of the square, and this produces the square on the right

$\blacksquare_1$	$\blacklozenge_1$	$\blacklozenge_4$	$\blacksquare_4$
$\blacktriangle_1$	$\bigstar_1$	$\bigstar_4$	$\blacktriangle_4$
$\blacktriangle_3$	$\bigstar_3$	$\bigstar_2$	$\blacktriangle_2$
$\blacksquare_3$	$\blacklozenge_3$	$\blacklozenge_2$	$\blacksquare_2$

$\longrightarrow$

$\blacksquare_1$	$\blacksquare_2$	$\blacksquare_3$	$\blacksquare_4$
$\bigstar_1$	$\bigstar_2$	$\bigstar_3$	$\bigstar_4$
$\blacktriangle_1$	$\blacktriangle_2$	$\blacktriangle_3$	$\blacktriangle_4$
$\blacklozenge_1$	$\blacklozenge_2$	$\blacklozenge_3$	$\blacklozenge_4$

1	6	13	2
10	5	14	9
7	12	11	16
8	3	4	15

$\longrightarrow$

1	15	8	2
5	11	12	14
10	16	7	9
6	4	3	13

Fig. 8 Transformation needed to produce the second magic square of subtraction of Figure 2

squares shown in Figure 2c and 2d. Application of the same permutation to both rows and columns preserves the magic property, so we are guaranteed that these are magic squares of subtraction too.

As clever as the above constructions are, they do not seem to be amenable to a generalization to higher orders. However, Kochański claims that he found a general method of construction of doubly even squares, of course without revealing any details. In spite of the lack of evidence, one might speculate that this general method possibly exploited the existence of squares of order 4 to build larger squares of doubly even orders. A similar method exists for squares of summation, and is sometimes known as the method of composite squares. This is a very obvious construction, so it seems probable that Kochański was aware of it. We will now show how this can be done.

Let us assume that $n = 4k$, $k \in \mathbb{N}$. The following proposition constructs magic square of subtraction of order n if a known square of order 4 is provided.

Proposition 2 *Let P be a magic square of subtraction of order 4, J be a matrix 4×4 with all entries equal to 1, and let Q be a block matrix $k \times k$ with entries $Q_{i,j} = P + (k(i-1) + j - 1)J$. Then Q is a magic square of subtraction of order $4k$.*

Matrix Q is simply a block matrix with entries consisting of sums of P and matrices $0, J, 2J \ldots, (k^2 - 1)J$, added in consecutive rows in increasing order. For example, for $k = 3$, it is given by

$$Q = \begin{pmatrix} P & P + 16J & P + 2 \cdot 16J \\ P + 3 \cdot 16J & P + 4 \cdot 16J & P + 5 \cdot 16J \\ P + 6 \cdot 16J & P + 7 \cdot 16J & P + 8 \cdot 16J \end{pmatrix}. \tag{16}$$

Note that each matrix $P + mJ$ is a magic square of subtraction, regardless of m, and its residuum is the same as the residuum of P.

Consider now, for example, the first row of Q, which consists of the first row of P followed by the first row of $P + 16J$ and then by the first row of $P + 32J$, which we will symbolically write as

$$Q_1 = (P_1, P_1 + 16J_1, P_1 + 32J_1). \tag{17}$$

Note that every entry of P_1 is smaller than every entry of $P + 16J$, and every entry of $P + 16J$ is smaller than every entry of $P + 32J$. For this reason,

$$\operatorname{res}(Q_1) = \operatorname{res}(P_1) + \operatorname{res}(P_1 + 16J_1) + \operatorname{res}(P_1 + 32J_1) \tag{18}$$

$$= \operatorname{res}(P_1) + \operatorname{res}(P_1) + \operatorname{res}(P_1) = 3 \operatorname{res}(P). \tag{19}$$

Exactly the same relationship holds for every row and column of Q, as well as for both diagonals, for the same reason as above. Obviously, this observation also easily generalizes to arbitrary k, when the residuum of every row, column, and both diagonals of Q would be equal to $k\,\mathrm{res}(P)$, proving that Q is indeed a magic square of subtraction. □

For $k = 2$, an example of construction of a square of order 8 using this method is shown below.

$$P = \begin{bmatrix} 1 & 6 & 13 & 2 \\ 10 & 5 & 14 & 9 \\ 7 & 12 & 11 & 16 \\ 8 & 3 & 4 & 15 \end{bmatrix}, Q = \left[\begin{array}{cccc:cccc} 1 & 6 & 13 & 2 & 17 & 22 & 29 & 18 \\ 10 & 5 & 14 & 9 & 26 & 21 & 30 & 25 \\ 7 & 12 & 11 & 16 & 23 & 28 & 27 & 32 \\ 8 & 3 & 4 & 15 & 24 & 19 & 20 & 31 \\ \hdashline 33 & 38 & 45 & 34 & 49 & 54 & 61 & 50 \\ 42 & 37 & 46 & 41 & 58 & 53 & 62 & 57 \\ 39 & 44 & 43 & 48 & 55 & 60 & 59 & 64 \\ 40 & 35 & 36 & 47 & 56 & 51 & 52 & 63 \end{array}\right]$$

This example uses as P one of the squares of order 4 constructed by Kochański, with residuum 8. The residuum of the resulting square Q is equal to $k\,\mathrm{res}(P) = 2{\cdot}8 = 16$.

What remains is the case of singly even squares, of order $n = 4k + 2$, $k \in \mathbb{N}$. Kochański apparently tried to construct the smallest order square of subtraction of this type, for $n = 6$, but he failed. For this reason, at the end of his paper, he challenged fellow mathematicians to find the square of order 6:

> *PROBLEMA I. In Quadrato Senarii, cellulas 36 complectente, numeros progressionis Arithmeticae, ab 1. ad 36. inclusive procedentes, ita disponere, ut subtractionis articio, prius explicato, in omnibus columnis, trabibus, & utraque Diagonali relinquatur numerus 18.*
>
> PROBLEM I. In a square of base 6, containing 36 cells, to arrange numbers of arithmetic sequence, proceeding inclusively from 1 to 36, in such a way that by the method of subtraction explained earlier, the number 18 remains in all columns, rows and both diagonals.

Although I was not able to discover any general method of construction of singly even squares, I found millions of $n = 6$ squares by brute-force computerized search. One such square is shown below. It has residuum 18, just as Kochański wished:

$$\begin{bmatrix} 1 & 11 & 12 & 13 & 29 & 2 \\ 17 & 5 & 19 & 20 & 8 & 33 \\ 34 & 14 & 6 & 9 & 28 & 23 \\ 32 & 16 & 10 & 7 & 24 & 25 \\ 35 & 26 & 27 & 22 & 21 & 15 \\ 3 & 36 & 18 & 31 & 30 & 4 \end{bmatrix}.$$

The question asked 330 years ago can, therefore, be considered answered, although the answer is not entirely satisfying. It has been obtained with the help of a machine, while Kochański wanted to *provocare homines ut vires suas experiantur*, that is, to *provoke people to put to the test their [mental] powers*. On the other hand, he spent a lot of time and effort designing and attempting to construct computing machines, so perhaps he would appreciate the fact that such machines were eventually constructed and that they helped to produce the example he asked for.

A general method for construction of singly even squares of orders higher than 6 remains, for now, an open problem. The known algorithms for construction of singly even magic squares of summation do not seem to provide any useful insight.

5 Enumeration

It is rather surprising to compare the number of existing magic squares of subtraction with the number of magic squares of summation. For magic squares of summation, their number is known up to order 5 (Trump 2012). It is customary to count not all magic squares, but only "distinct" ones, that is, the number of equivalence classes with respect to the group of 8 symmetries of the square (dihedral group D_4). Table 1 summarizes currently known enumeration results, counting only the above equivalence classes. In spite of the fact that magic squares of subtraction are subjected to strongly nonlinear constraints (equality of residua, which require sorting of vectors), for $n > 3$ they are apparently much more common than magic squares of summation, for which the constraints are linear. For $n = 4$, I found that the number of distinct magic squares of subtraction is 22,488, which is over 25 times more than the number of magic squares of summation of the same order. The exact number of squares of subtraction of order 5 remains unknown, but the non-exhaustive computer search shows that their number exceeds 10^9, possibly greatly, so it is certainly much larger than the number of magic squares of summation of order 5.

Direct computer enumeration of squares of order 5 is probably within the reach of current hardware, although I have not been able to complete it yet. Squares of order 6, on the other hand, are probably outside of the reach of today's computers, and may remain so for a long time (maybe even forever).

Table 1 Enumeration of magic squares of summation and subtraction

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
Number of distinct magic squares of summation	0	1	880	275,305,224	?
Number of distinct magic squares of subtraction	0	0	24,488	?	?

Acknowledgements The author acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of Discovery Grant. The computational part of this work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca) and Compute/Calcul Canada.

References

Bachet CG (1624) Problémes plaisans et délectables, qui se font par les nombres, 2nd edn. Pierre Rigaud & Associés, Lyon

Cantor M (1880) Vorlesungen über Geschichte der Mathematik. B. G. Teubner, Leipzig

Feingold M (ed) (2013) The New Science and Jesuit Science: Seventeenth Century Perspectives. Springer Netherlands

Fukś H (2012) Adam Adamandy Kochański's approximations of π: reconstruction of the algorithm. Mathematical Intelligencer 34:40–45, arXiv:1111.1739

Fukś H (2015) Three mathematical papers of Adam Adamandy Kochański – annotated English translation. Antiquitates Mathematicae 9:31–65

Günther S, Braunmühl A, Wieleitner H (1921) Geschichte der Mathematik. G. J. Göschen, Leipzig

Kircher A (1665) Arithmologia. Ex Typographia Varesij, Romae

Kochański AA (1682) Solutio theorematum ab illustri viro in actis hujus anni mense januario, pag. 28. propositorum. Acta Eruditorum pp 230–236

Kochański AA (1685) Observationes cyclometricae ad facilitandam praxin accomodatae. Acta Eruditorum 4:394–398

Kochański AA (1686) Considerationes quaedam circa quadrata & cubos magicos. Acta Eruditorum pp 391–395

Kochański AA (2003a) Opuscula. Cz. 1, prace zamieszczone w: "Miscellanea Curiosa Medico-Physica Academiae Naturae Curiosorum" i "Acta Eruditorum". WAM, Kraków, edited by B. Lisiak

Kochański AA (2003b) Opuscula. Cz. 2, prace zamieszczone w: Gaspar Schott, "Mathesis caesarea", Herbipoli 1662 ; Gaspar Schott, "Physica curiosa", Herbipoli 1662 ; Gaspar Schott, "Technica curiosa sive Mirabilia artis", Herbipoli 1664. WAM, Kraków, edited by B. Lisiak

Kochański AA (2003c) Opuscula. Cz. 3, prace zamieszczone w Gaspar Schott, "Cursus mathematicus", Herbipoli 1661. WAM, Kraków, edited by B. Lisiak

Lisiak B (2005) Adam Adamandy Kochański (1631–1700). Studium z dziejów filozofii i nauki w Polsce XVII wieku. Wydawnictwo WAM, Kraków, in Polish.

Lisiak B, Grzebień L (2005) Korespondencja Adama Adamandego Kochańskiego SJ (1657–1700). Wydawnictwo WAM, Kraków, in Latin with Polish summaries

Montucla JE (1754) Histoire Des Recherches Sur La Quadrature Du Cercle. Jombert, Paris

OEIS Foundation Inc (2015) The on-line encyclopedia of integer sequences. Http://oeis.org/A191642

O'Malley J, Bailey G, Harris S, Kennedy T (eds) (2016) The Jesuits: Cultures, Sciences, and the Arts, 1540–1773. University of Toronto Press

Pawlikowska Z (1969) Adam Adamandy Kochański i jego prace matematyczne. Rocznik Pol Tow Mat 11:19–49, in Polish.

Stasiewicz-Jasiukowa I (ed) (2004) Contributions of Jesuits to science and culture of the Polish-Lithuanian Commonwealth and under partitions. WAM, Kraków, in Polish

Trump W (2012) How many magic squares are there? Results of historical and computer enumeration. URL http://www.trump.de/magic-squares/howmany.html

Euler's E228: Primality Testing and Factoring via Sums of Squares

V. Frederick Rickey

The problem of distinguishing prime numbers from composite numbers and of resolving the latter into their prime factors is known to be one of the most important and useful in arithmetic

C. F. Gauss, *Disquisitiones Arithmeticae,* §329

Abstract How can you decide if a number is the sum of two squares? Euler began with the dumbest possible algorithm imaginable: Take the number, subtract a square, and check if the remainder is a square. If not, repeat, repeat, repeat. But Euler, being Euler, found a way of converting all those subtractions into additions. Then he did several things to speed up the computation even more. He applied this to 1,000,009, and—in less than a page—found that there are two ways to express this as a sum of squares. Hence, by earlier work in E228, it is not a prime. Amusingly, when he later described how to prepare a table of primes "ad millionem et ultra" (E467), he included this number as prime. So he then felt obliged to write another paper, E699 (1797), using another refinement of his method, to show that 1,000,009 is not prime.

Keywords Leonhard Euler • Number theory • Primes • Primality testing • Factoring • Sums of squares

V.F. Rickey (✉)
United States Military Academy, West Point, NY, USA
e-mail: fred.rickey@me.com

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*, Proceedings of the Canadian Society for History and Philosophy of Mathematics La Société Canadienne d'Histoire et de Philosophie des Mathématiques, DOI 10.1007/978-3-319-64551-3_7

1 Introduction

From the title of Euler's paper, "De numeris, qui sunt aggregata duorum quadratorum,"[1] one would think it only concerned expressing numbers as a sum of squares, a topic that occupied Euler for years. Of course this was part of the paper and we shall discuss the clever way that Euler did this, but our interest centers on the problem attacked in the final eight pages:

> To explore whether a given number of the form $4n + 1$ is prime or not.

We begin by describing how Christian Goldbach (1790–1764) got Leonhard Euler (1707–1783) interested in number theory and how Euler learned about the various claims that Fermat made. For our purposes the most important conjecture of Fermat that Euler learned was that every prime of the form $4n + 1$ has a unique splitting, i.e., it can be represented in one and only one way as the sum of two squares. For example, we know that $17 = 1^2 + 4^2$, $109 = 3^2 + 10^2$ and, to give an example of Euler's, $262\,657 = 129^2 + 496^2$. We shall not include a proof of this result because it is rather well known.[2]

Euler also proved that if a number can be split into two different ways as a sum of two squares, then it is composite. Moreover, he devised a simple algorithm for finding the factors of such composites. We shall provide several examples of how this works.

Finally we discuss a different technique that Euler used for showing that 1,000,009 is not a prime number.

2 Goldbach Sparks Euler's Interest in Number Theory

On May 27, 1727, three days after the twenty-year-old Euler arrived in Petersburg,[3] he met the 37-year- old Goldbach at a reception given by the Petersburg Academy's President Lorenz von Blumentrost [*EG Corr*, p. 17]. Goldbach had already published an elementary problem in number theory in his 1717 funeral ode for Leibniz: no square leaves remainder 2 when divided by 3 [*EG Corr*, pp. 5, 13, 15].

At the end of 1727, the court moved to Moscow for the coronation of Peter II (1715–1730), which was held at the Dormition Cathedral on 24 February 1728. In

---

[1] This paper, "On numbers which are the sum of two squares," was read to the Berlin Academy on March 20, 1749 and published in the *Novi commentarii academiae scientiarum Petropolitanae*, 4, 1758, pp. 3–40. We cite Euler's papers using their Eneström numbers. This one is E228.

[2] There are many proofs. See, e.g., Goldman 1998 (who gives three proofs), Wagon 1990, Zagier 1990, and Jameson 2010.

[3] Upon its founding in 1703, the official name of the town was "Sankt Pieter Burch." From 1724 to 1941 it was "Sanktpeterburg," but Euler and Goldbach used the Latin form "Petropoli" in their initial letters. For simplicity, we follow the editors of the *EG Corr* [p. 95] and use simply "Petersburg," with an infixical "s" that was never used in Russian.

January 1728, Goldbach left for Moscow to tutor Peter II and his four years older sister Natalia.

In October 1729, Daniel Bernoulli read [literally] to the Petersburg Academy a paper of Goldbach entitled "De terminis generalibus serierum." It was the most substantial of Goldbach's limited mathematical *œuvre* and dealt with interpolating terms in a sequence. This provided an opportunity for Euler to begin a correspondence with Goldbach that includes 196 letters dating from October 13 (24), 1729 until March 17, 1764, the year Goldbach died.[4] Although both were native German speakers, the correspondence began in Latin, but when Euler reported "Die *Geographie* ist mir *fatal*" in 1740 [*EG Corr*, n°34 August 21 (September 1), 1740],[5] they switched to a mixture of Latin (for the mathematics) and German (when no technical vocabulary was involved). Interestingly, they used different scripts for the different languages, a fact that is reproduced in the transcriptions of the letters in the *EG Corr*; see the example in §4 below.

In response to Euler's first letter, Goldbach added a postscript that profoundly influenced Euler's research:

> Do you know about Fermat's remark that all numbers of the form $2^{2^n} + 1$ (i.e., 3, 5, 17, and so on) are prime? He admitted however that he could not prove this, and as far as I know, nobody has proved it subsequently.[6] [*EG Corr*, n°2, (Nov. 20) Dec. 1, 1729]

Before long, Euler admitted to Goldbach that he had "started to ponder Fermat's theorem":

> Lately, reading Fermat's works, I came upon another rather elegant theorem stating that "any number is the sum of four squares" …
> There are other theorems in this book about the resolution of an arbitrary number into triangular and pentagonal numbers, into cubes, and so on; their demonstrations would indeed bring about a great progress in analysis. [*EG Corr*, n° 5, June 4, 1730]

---

[4]For biographical information about Goldbach and his relationship with Euler, see the *EG Corr*, pp. 3–31.

[5]This is a letter number 34 in the new Euler-Goldbach correspondence, *Leonhardi Euleri commercium epistolicum cum Christians Goldbach*, volume IV.A.4 of Euler's *Opera Omnia*, which is in two parts. The first has the correspondence in the original languages, the second has English translations. There is also a useful introduction and an abundance of explanatory notes. The work is edited by Franz Lemmermeyer and Martin Mattmüller and was published by Springer in 2015. These volumes are designated *EG Corr* here. Sometimes when a letter is cited by number, we omit "*EG Corr*."

[6]In E26 (1738) Euler "observed after thinking about this for many days" that the fifth Fermat number, $2^{2^5} + 1 = 4\,294\,967\,297$, is not prime because it is divisible by 641. Although he does not give any clue about how he discovered this fact he noted that it can "be seen at once by anyone who cares to check."

In E134 (1750), §32, Euler noted that the only trial divisors that he needed to consider were of the form $64n + 1$. See Sandifer 2015.

3 Euler Learns of Fermat's Work

Unfortunately, Euler did not report the title of "Fermat's works" or "this book." So what did he read? Let us consider the possibilities.

3.1 Diophantus 1670

Did Euler have access to Samuel de Fermat's edition of *Diophanti Alexandrini arithmeticorum libri sex*? Pierre Fermat, the father of Samuel, had a copy of the 1621 edition of this work which had been translated and edited by Bachet. The senior Fermat wrote annotations in the 1621 edition which were first published in the 1670 edition. Euler did have access to this 1670 work, but perhaps not early on, as Weil suggests (1984, p. 173), for he does not mention it in the *EG Corr* until February 13, 1748, n° 125, when he paraphrased what has come to be known as Fermat's Last Theorem and in n° 128, (May 28) June 8, 1748, where he quoted from the work. Had he read it earlier, it would be surprising that he did not mention it.

While it is doubtful that Commentary 7, as Pierre de Fermat's annotation is now denoted, is the source of Euler's interest in sums of squares, it is the one which interests us here[7]:

> Nvmerus primus qui superat vnitate quaternarii multiplicem semel tantùm est hypotenusa trianguli rectanguli, eius quadratus bis, cubus 3. quadratoquadratus 4 &c. in infinitum. [Diophantus 1670, p. 127]
>
> A prime number of the form $4n + 1$ is the hypotenuse of a right-angled triangle in one way only, its square is so in two ways, the cube in three, its biquadrate in four ways, and so on *ad infinitum*. [Heath 1910, p. 106]

3.2 Wallis 1658

There is no doubt that Euler consulted the *Commercium epistolicum* of John Wallis (1616–1703), for he cited it by title in the very paper, E26, where he noted that $2^{2^5} + 1$ was divisible by 641. In the last[8] letter in the volume, which is from Fermat to Kenelm Digby, June 1658, Fermat wrote:

---

[7]The 48 "Observatio Domino Petri de Fermat" are scattered through the 1770 Diophantus, where they occur in print for the first time. The Latin originals have been reprinted in the *Œuvers de Fermat*, volume 1 (1891), pp. 291–342. French translations are in volume 3 (1896), pp. 241–274.

[8]Euler is a bit confused here for he referred to this as the penultimate letter. The penultimate letter, XLVI, is the note, in English dated June 19, 1658, whereby Digby conveys Fermat's letter to Wallis further requesting that he pass it on to Brouncker.

> Omnis numerus primus qui unitate superat quaternarii multiplicem est composites ex duobus quadrati. Huiusmodi sunt 5.13.17.29.37.41. &c. [Wallis 1658, p. 186; reprinted *Œuvers de Fermat*, II, letter XCVI, p. 403]
> All prime numbers which exceed a multiple four by one are composed of two squares. Such are 5, 13, 17, 29, 37, 41, etc.

Immediately before this statement Fermat declared that he had a most rigorous demonstration (firmissimis demonstrationibus), but like almost all of Fermat's claims, the proof has not been found. So this is now the challenge that Euler will deal with.

This letter from Fermat to Digby was reprinted in the *Operum mathematicorum volume tertium* (1699) of Wallis. Euler cited this work in a different context in n^o 53, July (19) 30, 1742.

3.3 Frénicle de Bessy 1676

The *Traité des triangles rectangles en nombres* of Frénicle de Bessy was mentioned by Euler in n^o 98, January 25, 1746. This work contains Fermat's proof by descent that there is no triangle with integral sides whose area is a square. The first time that Fermat made this claim, in a letter to Frènicle dated June 15, 1641, he dubbed it the fundamental proposition of rectangular triangles.

> Tout nombre premier, qui surpasse de l'unité un multiple du quaternaire, est une seule fois la somme de deux quarrés, et une seule fois l'hypoténuse d'un triangle rectangle. [See *Œuvers de Fermat*, II, letter XLV, December 25, 1640, p. 213].

Note that this statement of Fermat is more precise: A prime of the form $4n + 1$ can be represented as a sum of two squares in one *and only one* way.

3.4 Fermat 1679

The only book of Fermat's that Euler had access to was his posthumous *Varia opera mathematica*. Calinger 2016, p. 79, writes, without giving a reference, that Euler read this in 1729. It contains Fermat's letter to Roberval of August 1640 and a letter to Digby of June 1658.

4 Euler Publishes His Results

While the above information is not conclusive, it is our best information about how Euler could have learned about "Fermat's Theorem," as Euler often called it, that every prime of the form $4n + 1$ is always the sum of two squares. For seven years, Euler worked hard to prove this result and finally he was able to write Goldbach on May 6, 1747, n° 115, that he has found, "after a lengthy effort" a proof of this result.

He presented his results on March 20, 1749 at a meeting of the Berlin Academy, even though he knew there was a gap in Proposition V (every prime number which exceeds a multiple of four by one is a sum of two squares), so he was only able to give an "Attempt at a Proof" (Tentamen demonstrationis), a blemish that persisted in the published paper, E228.

His ideas were in the right direction and less than a month after the meeting where Euler presented his results to the Berlin Academy, Euler filled the gap. On April 12, 1749, n°138, he wrote to Goldbach:

> Nunmehro habe ich endlich einen bündigen Beweiß gefunden daß ein jeglicher *numerus primus* von dieser *Form* $4n+1$ eine *summa duorum quadratorum* ist. Es sey $\boxed{2}$ das Zeichen der Zahlen, welche *summa duorum quadratorum* sind; so sind meine Sätze folgende:
> I. *Si* $a = \boxed{2}$ *et* $b = \boxed{2}$ *erit etiam* $ab = \boxed{2}$, wovon der Beweiß leicht.
> Now I have finally discovered a conclusive proof that every prime number of the form $4n+1$ is a sum of two squares. Letting the sign $\boxed{2}$ denote those numbers which are sums of two squares, my propositions are as follows:
> 1. If $a = \boxed{2}$ and $b = \boxed{2}$, then $ab = \boxed{2}$; this is easy to prove. [*EG Corr*, pp. 446 and 992].

Euler presented this correct proof in E241: "Demonstratio theorematis Fermatiani omnem numerum primum formae $4n+1$ esse summam duorum quadratorum," which was published only in 1760 [n° 138, note 2].

5 Expressing Integers as Sums of Squares

One way to decide if a number such as 1,000,009 is prime is by dividing by all primes less than 1000. There are 168 such primes so it seems unlikely that Euler, even given his prodigious computational ability, would use this method. Euler had a better way. He knew that a number of the form $4n + 1$ is prime if and only if it could be represented as a sum of two square in one and only one way. To begin, we present Euler's computation.

The procedure that Euler[9] used to decide if a number, N, is the sum of two squares is just what one would expect. Subtract the smallest square less than N and check if the difference is a square. If it is, then N is the sum of two squares. If not, then subtract the next smallest square, p^2, and see if $N - p^2$ is a square. Repeat until the difference is smaller than $N/2$ (there is no reason to continue for if the difference is a square it will already have been noted). The number of times that this question

[9]Fermat had a factorization method that bears some similarities with Euler's, but Euler could not have known about it because it was not published until the nineteenth century by Charles Henry, a librarian, history of mathematics specialist, and co-editor of the *Œuvers de Fermat*. For a translation and explanation, see Chabert 1999, pp. 264–266. Curiously, Fermat's method is better known today than Euler's. See Ore 1948, pp. 56–58 and Burton 1976, pp. 94–97.

is answered affirmatively is the number of times that N can be written as a sum of squares. If this number is 1, then N is prime; otherwise it is composite.

While conceptually simple, this procedure involves a tedious computation. Let $L = \lfloor\sqrt{N}\rfloor$ be the largest integer whose square is less than N. Then we must compute the differences

$$N - L^2, \quad N - (L-1)^2, \quad N - (L-2)^2, \quad \ldots \quad N - (L-a)^2,$$

where $N - (L-a)^2$ is just less than $N/2$.

In Euler's Example 1, $N = 82421$ and $L = 283$ and so it would take about 140 subtractions to complete the process. Euler probably would not have written this paper had he not found a clever and simpler way to do the problem. His trick was to replace subtraction by addition.

Let us consider a small example: is 12037 a prime? Start subtracting squares and see if the difference is a square. Since $\lfloor\sqrt{12037}\rfloor = 109$,

- First subtract 109^2 from 12037.
- Then subtract 108^2 from 12037.
- Then subtract 107^2 from 12037.
- Then subtract 106^2 from 12037.

If one and only if one of these differences is a square, great, we are done. Our number is a prime. But if more than one square (or none) shows up, our number is composite. This technique does not seem promising, but let us try.

$$\begin{array}{r} 12037 \\ 109^2 = 11881 \\ \hline 156 \end{array} \quad \begin{array}{r} 12037 \\ 108^2 = 11664 \\ \hline 373 \end{array} \quad \begin{array}{r} 12037 \\ 107^2 = 11449 \\ \hline 588 \end{array} \quad \begin{array}{r} 12037 \\ 106^2 = 11236 \\ \hline 801 \end{array}$$

$$\begin{array}{r} 12037 \\ 105^2 = 11025 \\ \hline 1012 \end{array} \quad \begin{array}{r} 12037 \\ 104^2 = 10816 \\ \hline 1221 \end{array} \quad \begin{array}{r} 12037 \\ 103^2 = 10609 \\ \hline 1428 \end{array} \quad \begin{array}{r} 12037 \\ 102^2 = 10404 \\ \hline 1633 \end{array}$$

$$\begin{array}{r} 12037 \\ 101^2 = 10201 \\ \hline 1836 \end{array} \quad \begin{array}{r} 12037 \\ 100^2 = 10000 \\ \hline 2037 \end{array}$$

Is it really necessary to do all of these subtractions? Can we eliminate some of them? Since we have been subtracting squares, we note that their last digits are 0, 1, 4, 5, 6, or 9. When we subtract these from 12037, the last digits are 7, 6, 3, 2, 1, or 8, respectively. But only those ending in 1 and 6 can be squares. So the number we squared must end in 9, 6, 4, 1. Consequently, we only need to consider the following numbers when we subtract from 12037:

$$\begin{array}{cccc} 109^2 & 106^2 & 104^2 & 101^2 \\ \\ 99^2 & 96^2 & 94^2 & 91^2 \\ \\ 89^2 & 86^2 & 84^2 & 81^2 \\ & \vdots & & \end{array}$$

Entries in the first column have the form $(109 - 10k)^2$, for $k = 0, 1, 3, \ldots$. For the moment let us just consider this column.

Now we shall enjoy a touch of Euler's cleverness.

Suppose N is the number to be tested and $p = \lfloor\sqrt{N}\rfloor$ is the largest square less than N. Now compute $N - p^2$. The number below it in the column is $N - (p - 10)^2$. But $N - (p - 10)^2 = N - p^2 + 20p - 100$. So if we ADD $20p - 100$ to $N - p^2$ we obtain $N - (p - 10)^2$, the second entry in the column. Now iterate

Number to test	N		
Subtract	p^2		
	$N - p^2$		
plus	$20p - 100$		
	$N - p^2 + 20p - 100$	$=$	$N - (p - 10)^2$
plus	$20p - 300$		
	$N - p^2 + 40p - 400$	$=$	$N - (p - 20)^2$
plus	$20p - 500$		
	$N - p^2 + 60p - 900$	$=$	$N - (p - 30)^2$

In this computation only one subtraction $(N - p^2)$ and one easy multiplication and subtraction $(20p - 100)$ were performed. All the rest of the computation is addition.

Nota bene: Subtraction has been replaced by addition.

Euler even had tables of primes[10] and squares[11] to simplify his life.

[10]We know from Euler's unpublished *Catalogus librorum meorum* in his "Notebook VI" that he owned a copy of Johann Gottlob Krüger's *Gedancken von der Algebra, nebst den Primzahlen von 1 biß 10000000* (1746) [*EG Corr*, p. 1071, n. 7]. The title has a misprint; the last prime listed is 100,999. Euler had a copy of this at least by 1760, for he alluded to it in E283 (1764). Even earlier, Euler wrote to Goldbach (April 18 (29), 1741, n° 36) that he planned to present him with a copy of Frans van Schooten's *Exercitationum mathematicarum libri quinque* (1657) which contains, pp. 393–403, a list of the prime numbers up to 10,000: *Syllabus numerorum primorum, qui continentur in decem prioribus chiliadibus*. Euler notes that he has copied out the prime numbers of the form $4n + 1$ up to 3000. It is not known if Euler had access to a factor table; for history see Bullynck 2010.

[11]While I have no information that Euler had access to it, a table of the squares of all numbers up to 100,000 had been published, viz.: Johann-Hiob Ludolf, *Tetragonometria tabularia, Quâ Per Tabulas Quadratorum a Radice Quadrata I. Usque Ad I00000*, Leipzig: Groschian, 1690.

Now we present the entire computation.

	p	12037	p	12037	p	12037	p	12037
	109	11881	106	11236	104	10816	101	10201
		156		801		1221		1836
20p-100		2080		2020		1980		1920
	99	2236		2821		3201		3756
20p-300		1880		1820		1780		1720
	89	4116		4641		4981		□ 5476
20p-500		1680		1620		1580		1520
	79	5796		6261				

The number with the square (□) is $12037 - 81^2$ which has square root 74. Thus

$$12037 = 81^2 + 74^2,$$

and this is the only square, so 12037 is a prime.

Here is another example: Is 12077 prime?

	p	12077	p	12077	p	12077	p	12077
	109	11881	106	11236	104	10816	101	10201
		□ 196		□ 841		1261		1876
20p-100		2080		2020		1980		1920
	99	2276		2861		3241		3796
20p-300		1880		1820		1780		1720
	89	4156		4681		5021		5516
20p-500		1680		1620		1580		1520
	79	5836		6301		6601		7036
20p-700		1480		1420		1380		1320
	69	7316		7721		7981		8356

There are two squares here, so $14^2 + 109^2 = 29^2 + 106^2 = 12077$ and thus 12077 is composite. We could have stopped after the fourth row, but wanted to see if there other ways that 12077 could be split.

The arrangement of the computation is beneficial. There are patterns in Euler's tables that help detect errors.

[other editions were published in 1709 and 1712]. See Denis Roegel, A reconstruction of Ludolf's *Tetragonometria tabularia* (1690), Technical report, LORIA, Nancy, 2013. Even if Euler did not have such a table, he certainly knew the 22 possible last two digits for a square.

We have given several examples above where the number tested ended in 7. Note that in those cases there are 4 columns. What about numbers that end in 1, 3, or 9? By giving an analysis such as we gave for numbers ending in 7, we can determine

if the given number ends in	the squares to be subtracted end in	and the roots of these squares end in
1	0, 1, 5, 6	0, 1, 4, 5, 6, 9
3	4, 9	2, 3, 7, 8
7	1, 6	1, 4, 6, 9
9	0, 4, 5, 9	0, 2, 3, 5, 7, 8

6 A Factoring Formula

On February 16, 1745, after remarking that Goldbach's ten-year-old Godson, Johann Albrecht Euler (1734–1800), was in better health and was learning quadratic equations, Euler added two results of interest:

> Moreover it can also easily be shown that if some number can be split into two squares in two ways, it is not prime. For let $N = a^2 + b^2 = c^2 + d^2$; then
>
> $$N = \frac{\left((a-c)^2 + (b-d)^2\right)\left((a+c)^2 + (b-d)^2\right)}{4(b-d)^2}. \tag{1}$$
>
> Furthermore the following theorem is also correct: If the number $4n + 1$ can be split into two squares in a unique way, then it will certainly be prime [*EG Corr*, n°87, February 16, 1745]

Euler did not publish the above formula in E228 (Did he ever publish it?). Perhaps that is because it is a *deus ex machina* giving no insight into how the result might be derived.

The second of these results is not quite correct, but by the time E228 was published, Euler added the condition that the two squares must be "prime between themselves" to Proposition VI (§35) and pointed out its importance in Example 6 (§51). Note 6 of page 857 of the *EG Corr* gives the impression that this error was not corrected for more than two centuries.

Proposition VII in §40 of E228 gave a derivation of a related result: "A number which can be written in two or more different ways as a sum of two squares is not prime." Note that Euler does not assume that N has the form $4n + 1$ in his proof:

Let $N = a^2 + b^2 = c^2 + d^2$. Without loss of generality, $a > b$ and $c > d$. The splittings are different, so $a \neq c$ and $b \neq d$. If $a > c$, then $b < d$, so there are integers x and y with

$$a = c + x \qquad d = b + y$$

From $a^2 + b^2 = c^2 + d^2$, he obtained $2cx + x^2 = 2by + y^2$.

Because the left-hand side of this equation is divisible by x and the right side by y, both sides are equal to xyz for some z. Thus

$$c = \frac{yz - x}{2}, \quad b = \frac{xz - y}{2}, \quad \text{and} \quad a = \frac{yz + x}{2}. \tag{2}$$

Then, with a bit of algebra, we have

$$N = a^2 + b^2 = \frac{y^2z^2 + 2xyz + x^2 + x^2z^2 - 2xyz + y^2}{4} = \frac{(x^2 + y^2)(1 + z^2)}{4}. \tag{3}$$

It is tempting to think that the integer $x^2 + y^2$ is a factor of N, but that is not necessarily the case, for we do not know that $(1 + z^2)/4$ is an integer. Euler notes that if $x^2 + y^2$ is composite, "some factors of it will be a divisor of N." Perhaps an example will clarify this.

On June 30, 1741, n°52, Euler used his earlier result that $2^{2^5} + 1 = 641 \cdot 6\,700\,417$, to find the double splitting:

$$2^{2^5} + 1 = 65\,536^2 + 1^2 = 62\,264^2 + 20\,449^2.$$

Using the notation from the above proof, we have

a = c + x	d = b + y
65536 = 62264 +x	20449 = 1 + y
x = 3272	y = 20448

Thus

$$x^2 + y^2 = 3272^2 + 20448^2 = 428\,826\,688.$$

This cannot be a factor of $2^{2^5} + 1 = 4\,294\,967\,297$ for an even number cannot be a factor of an odd number. However, if we divide this by 64, we obtain the larger factor of $2^{2^5} + 1$.

Still more explanation is needed. From

$$c = \frac{yz - x}{2} \quad \text{we obtain} \quad z = \frac{a + c}{d - b}$$

and thus

$$z = \frac{65536 + 62642}{20449 - 1} = \frac{25}{4}$$

so

$$1 + z^2 = 1 + \left(\frac{25}{4}\right)^2 = \frac{641}{16}$$

and hence

$$N = a^2 + b^2 = \frac{(x^2 + y^2)(1 + z^2)}{4} = \frac{428\,826\,688 \cdot 641}{64} = 6\,700\,417 \cdot 641.$$

Finally if we substitute the values of x, y, and z in terms of a, b, c, and d in the right-hand side of equation (3) we obtain (1). So it is not as mysterious as it first appeared.

7 Finding the Factors

In §41, Euler gave an algorithm for finding the factors of a number which has two splittings. First we give his argument and then we present a simpler way to obtain the factors, a technique which he certainly knew. Suppose

$$N = a^2 + b^2 = c^2 + d^2$$

Then by Proposition IV of E228, which stated that any number which divides $a^2 + b^2$, with a and b relatively prime, must also be a sum of squares. Thus

$$N = (p^2 + q^2)(r^2 + s^2).$$

Then Euler stated something that is true but not obvious: "these will certainly be $a = pr + qs$, $b = ps - qr$, and $c = ps + qr$, $d = pr - qs$." Hence "from this is obtained $a - d = 2qs$ and $c - b = 2qr$, so"

$$\frac{r}{s} = \frac{c - b}{a - d}$$

Euler next reduced this fraction to lowest terms and called the result $\frac{r}{s}$, a poor choice of notation for now r and s have two meanings. Instead, we take $k = gcd(c{-}b, a{-}d)$. Then $c - b = kR$ and $a - d = kS$ for some relatively prime integers R and S. Thus we have

$$\frac{r}{s} = \frac{c - b}{a - d} = \frac{kR}{kS} = \frac{R}{S}.$$

Euler then noted, with our change of notation, that $R^2 + S^2$ is a divisor of N, "unless it is even, for if it is even, then it should be assumed to be half of this."

We could have computed this way:

$$r^2 + s^2 = k^2(R^2 + S^2)$$

and so

$$N = (p^2 + q^2)k^2(R^2 + S^2).$$

The following seems to be the way that Euler used this in practice. From $a^2 + b^2 = c^2 + d^2$ we obtain

$$(a + c)(a - c) = (d + b)(d - b)$$

and hence

$$\frac{a+c}{d+b} = \frac{d-b}{a-c} \quad \text{and} \quad \frac{a+c}{d-b} = \frac{d+b}{a-c}.$$

We could also have obtained

$$(a + d)(a - d) = (c + b)(c - b)$$

and hence

$$\frac{a+d}{c+b} = \frac{c-b}{a-d} \quad \text{and} \quad \frac{a+d}{c-b} = \frac{c+b}{a-d}.$$

Thus to obtain the factors of N, we need to reduce the fractions on the left-hand sides of these equations:

$$\frac{a+c}{d+b}, \quad \frac{a+c}{d-b}, \quad \frac{a+d}{c+b}, \quad \text{and} \quad \frac{a+d}{c-b}$$

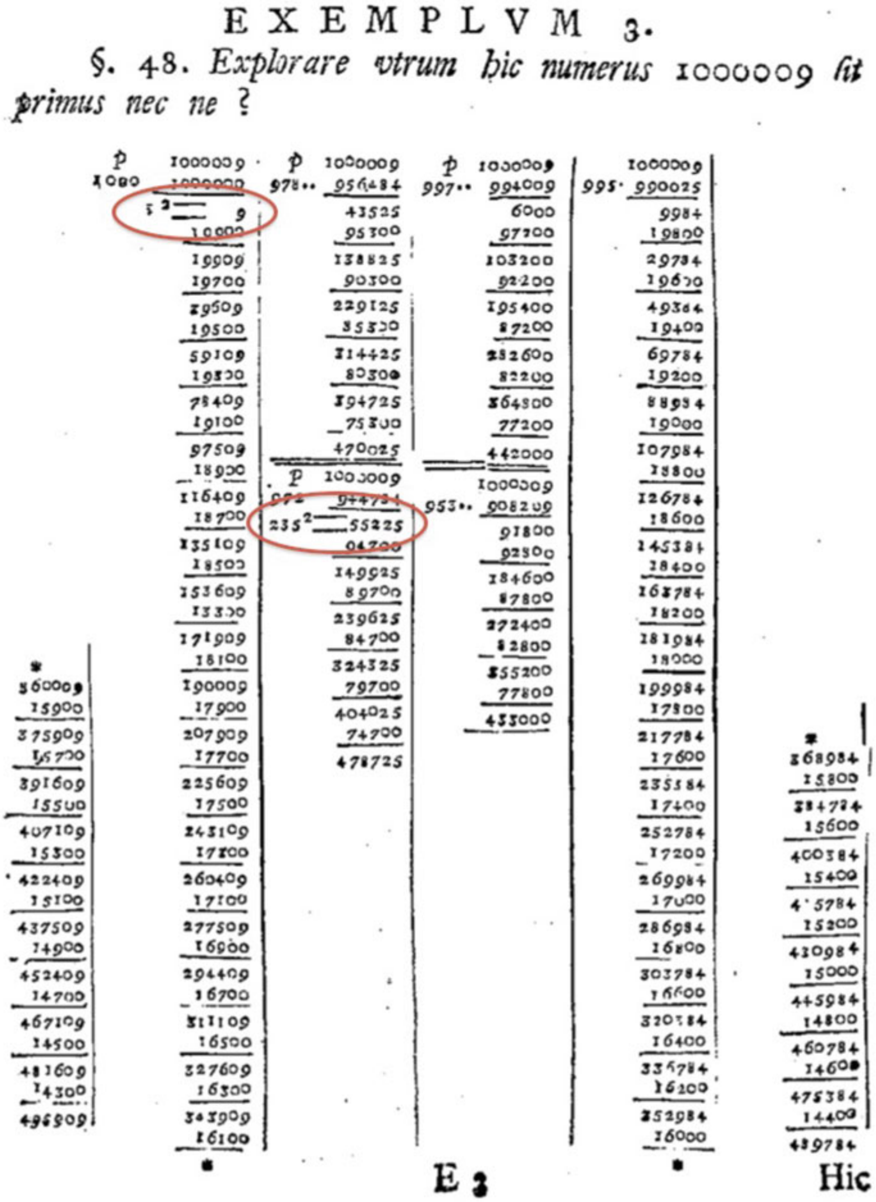
to obtain the values of R and S that we need.

Some examples will make this algorithm clearer.

8 Some Examples

Euler gave six examples, all of the form $4n + 1$. Of these, 82421, 100981, and 262657, each have one splitting and so are prime. The number 233033 cannot be split as the sum of two squares so is composite. However, $32129 = 95^2 + 152^2$ has precisely one splitting but 95 and 152 are not relatively prime, having greatest common divisor 19, so 32129 is composite, in fact, $32129 = 19^2 \cdot 89$.

(1) We now consider the third example in E228: to factor 1000009. At the time, for primality testing, large was over 1,000,000 because that was as far as the tables of primes went, so this would have been an impressive example [*EG Corr*, n°163, October 28, 1752]. Here is Euler's computation:



EXEMPLVM 3.

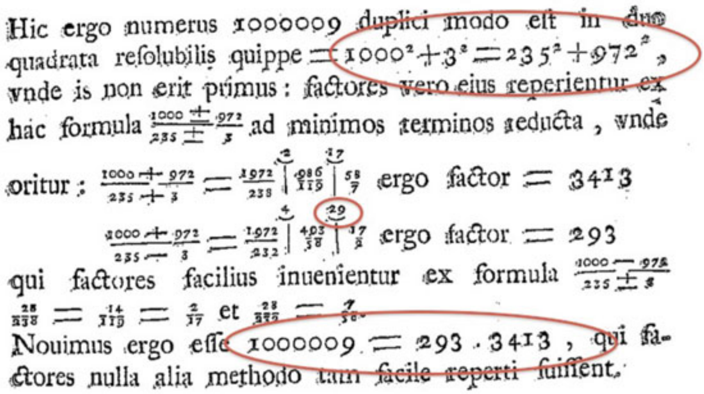
§. 48. *Explorare utrum hic numerus* 1000009 *sit primus nec ne?*

	p 1000009	p 1000009	p 1000009	1000009	
	1000 1000000	978.. 956484	997.. 994009	995· 990025	
	3² 9	43525	6000	9984	
	19900	95300	97200	19800	
	19909	138825	103200	29784	
	19700	90300	92200	19600	
	39609	229125	195400	49384	
	19500	85300	87200	19400	
	59109	314425	282600	69784	
	19300	80300	82200	19200	
	78409	394725	364800	88984	
	19100	75300	77200	19000	
	97509	470025	442000	107984	
	18900	p 1000009	1000009	18800	
	116409	972 944784	953.. 908209	126784	
	18700	235² 55225	91800	18600	
	135109	94700	92800	145384	
	18500	149925	184600	18400	
	153609	89700	87800	163784	
	18300	239625	272400	18200	
	171909	84700	82800	181984	
*	18100	324325	355200	18000	
360009	190009	79700	77800	199984	
15900	17900	404025	433000	17800	
375909	207909	74700		217784	*
15700	17700	478725		17600	368984
391609	225609			235384	15800
15500	17500			17400	384784
407109	243109			252784	15600
15300	17300			17200	400384
422409	260409			269984	15400
15100	17100			17000	415784
437509	277509			286984	15200
14900	16900			16800	430984
452409	294409			303784	15000
14700	16700			16600	445984
467109	311109			320384	14800
14500	16500			16400	460784
481609	327609			336784	14600
14300	16300			16200	475384
495909	343909			352984	14400
	16100			16000	489784
	*			*	

E 3 Hic

The number 1000009 ends in 9, so according to the table at the end of §5, Euler's Example 3 should have 6 columns and, indeed it does, but we must look carefully. The first and sixth columns are actually continuations of the second and fifth. The middle two columns actually count for two columns each, witness the "*p* 1000009" that occurs twice in each of the middle columns.

Here is Euler's conclusion:


Hic ergo numerus 1000009 duplici modo eſt in duo quadrata reſolubilis quippe $= 1000^2 + 3^2 = 235^2 + 972^2$, vnde is non erit primus: factores vero eius reperientur ex hac formula $\frac{1000 \pm 972}{235 \pm 3}$ ad minimos terminos reducta, vnde oritur: $\frac{1000+972}{235+3} = \frac{1972}{238} \,\big|\, \frac{986}{119} \,\big|\, \frac{58}{7}$ ergo factor $= 3413$

$\frac{1000+972}{235-3} = \frac{1972}{232} \,\big|\, \frac{493}{58} \,\big|\, \frac{17}{2}$ ergo factor $= 293$

qui factores facilius inueniuntur ex formula $\frac{1000-972}{235 \pm 3}$ $\frac{28}{238} = \frac{14}{119} = \frac{2}{17}$ et $\frac{28}{232} = \frac{7}{58}$. Nouimus ergo eſſe $1000009 = 293 \cdot 3413$, qui factores nulla alia methodo tam facile reperti fuiſſent.

After he has found that it has two splittings Euler noted that

$$\frac{1000+972}{235+3} = \frac{58}{7} \quad \text{and} \quad \frac{1000+972}{235-3} = \frac{17}{3}$$

and thus the factors of 1 000 009 are

$$58^2 + 7^2 = 3413 \quad \text{and} \quad 17^2 + 3^2 = 293.$$

Euler concluded, with a bit of satisfaction, that these factors could not be "so easily discovered by any other method." [§48]

(2) Since Euler was fond of doing problems several ways, we return to the example in the previous section. Like Euler, we are using the time honored research method of checking new techniques on problems where one already knows the answer. By factoring the double splitting:

$$2^{2^5} + 1 = 65\,536^2 + 1^2 = 62\,264^2 + 20\,449^2$$

as

$$(65536 - 62264)(65536 + 62264) = (20449 - 1)(20449 + 1)$$

or

$$3272 \cdot 127\,800 = 20\,448 \cdot 20\,450.$$

we obtain

$$\frac{20\,448}{127\,800} = \frac{4}{25} \quad \text{and} \quad \frac{20\,448}{3\,272} = \frac{2556}{409},$$

where we have reduced the fractions to lowest terms and note that the numerator and denominator of each are relatively prime. From this, we obtain the factors

$$4^2 + 25^2 = 641 \quad \text{and} \quad 2\,556^2 + 409^2 = 6\,700\,417.$$

If we had factored the double splitting as

$$(65536 - 20449)(65536 - 20449) = (65536 - 1)(65536 + 1)$$

we would have obtained $21^2 + 29^2 = 1\,282$ and $2\,147^2 + 2\,965^2 = 13\,392\,250$, each of which is even so we need to divide by two, thus obtaining the same result.

It seems that Euler did not know that 6 700 417 was prime. He could have discovered this by using his techniques to write it as a sum of squares in a unique way, but that would be a long computation.

(3) The only other example that Euler did in E228 (§43) is a classic one, dating back to Diophantus,[12] Fibonacci, and ancient India. From

$$85 = 9^2 + 2^2 = 7^2 + 6^2$$

we obtain four fractions:

$$\frac{9+7}{6+2} = \frac{2}{1}, \quad \frac{9+7}{6-2} = \frac{4}{1}, \quad \frac{9+6}{7+2} = \frac{5}{3}, \quad \text{and} \quad \frac{9+6}{7-2} = \frac{3}{1}.$$

The first two of these give rise to the factors $2^2 + 1^2 = 5$ and $4^2 + 1^2 = 17$, so $85 = 5^2 + 17^2$. The second two generate $5^2 + 3^3 = 34$ and $3^2 + 1^2 + 1^2 = 10$, both of which are even, so we must divide them by 2 to obtain the factors.

9 E699: Is 1,000,009 a Prime Number?

Euler made an error. His Example 3 from E228 showed that 1000009 can be split two ways as a sum of two squares, and hence is composite. But in E467 where Euler discusses how one might go about making a large table of primes, this number is

[12]The editors of the *EG Corr*, p. 40, note that Diophantus knew how, from the factors 5 and 13, to represent 65 as the sum of two squares.

listed as a prime. Now the computation in E228 is correct, so Euler could have just pointed out the error; there is no need for a second paper. But Euler wrote another paper because he found a new way to conclude that 1000009 is not prime.

Clearly $1000009 = 1000^2 + 3^2$. But is there another way to write it as a sum of squares? Suppose $1000009 \equiv x^2 + y^2 \equiv 9 \pmod{10}$. Then there are the following possibilities, where the only values modulo 10 for x^2 are listed in the first column.

x^2	y^2	
0	9	Known solution
1	8	8 is not the last digit of a square
4	5	Look here for a solution
5	4	Look here for a solution
6	3	3 is not the last digit of a square
9	0	Known solution

Hence $1000009 - x^2$ is divisible by 5, and thus by 25. Let $x = 25a + 3$ (the 3 eliminates the 9), then

$$1000009 - x^2 = 1000009 - (25a + 3)^2 = 1000000 - 6 \cdot 25a - 25^2a^2$$

When we divide by 25, we see that $40000 - 6a - 25a^2$ must be a square. Then there are three cases:

Case A If a is even, say $a = 2b$ then, after dividing by 4, we obtain

$$A = 10000 - 3b - 25b^2.$$

Case B If a is odd and $a = 4c + 1$, then

$$B = 39969 - 224c - 400c^2.$$

Case C If a is odd and $a = 4d - 1$, then

$$C = 39981 + 176d - 400d^2.$$

In Case B we want to know if $B = 39969 - 224c - 400c^2$ is a square? Now c can be positive or negative. We list the positive values of c in the left three columns and the negative values in the right three.

c	$400cc - 224c$	Diff.	c	$400cc + 224c$	Diff
0	0		0	0	
		176			624
1	176		1	624	
		976			1424
2	1152		2	2048	
		1776			2224
3	2928		3	4272	
		2576			3024
4	5504		4	7296	

Here is Euler's computation for Case B:

39969	39969	31089	28849
177	624	4176	4624
39793	39345	26913	24225
976	1424	4976	542.
38817	37921	21937	18801
1776	2224	5776	6224
37041	35697	16161	12577
2576	3024	6576	7024
34465	32673	9585	5553
3376	3824	7376	
31089	28849	* 2209	

The asterisk in the third column[13] indicated that 2209 was a square. In fact $2209 = 47^2$. This square arose from $c = -10$, so $a = 4c+1 = 4(-10)+1 = -39$. Earlier Euler let $x = 25a + 3$, so $x = 25(-39) + 3 = -972$. And hence $1000009 - xx = 55225 = 235^2$ So we have a double splitting: $1000009 = 1000^2 + 3^2 = 972^2 + 235^2$. Thus 1000009 is *not* prime, contrary to E467.

From the double splitting, we have $1000^2 - 235^2 = 972^2 - 3^2$, i.e., $(1000 + 235)(1000 - 235) = (972 + 3)(972 - 3)$, or $1235 \cdot 765 = 975 \cdot 969$. Hence

$$\frac{1235}{975} = \frac{19}{15} \quad \text{and} \quad \frac{1235}{969} = \frac{65}{51}.$$

We have $19^2 + 15^2 = 293$ and $65^2 + 51^2 = 6826$. Because the later number is even we need to take half of it to obtain a divisor. Thus $1000009 = 293 \cdot 3413$.

[13] In the first column 177 is a typo for 176.

Because case B determined that 100009 is composite, there is no reason to consider cases A and C. However, Euler does discuss them, for the methods involved are slightly different. Presumably he hoped that those ideas would be useful later.

10 Conclusion

It would be wonderful to report that this method of Euler had a great impact on factorization, but, sadly, we have found only one work where it was used and it is contemporary with Euler's work, namely Georg Wolfgang Krafft, "De divisoribus numerorum indagandis," *Novi commentarii academiae scientiarum Petropolitanae*, 3 (1751/1753), pp. 109–124. In the nineteenth century, only one relevant article has been located, namely Wertheim (1887) and it is a textbook exposition of Euler's method.

Euler used variants of his method in a number of papers—for example, E241, E256, E369, E461, E566, E718, E792, and E798—but a discussion of them must be left for another time.

Of course there are many authors who took up the problem of factoring such as Gauss in his *Disquisitiones Arithmeticae* of 1801. As progress is made in mathematics, older methods, interesting as they are, are often forgotten. But in this case, there has been a recent explosion of interest in factoring, and Euler's method is being used in current research. See, for example, Riesel 1985 and McKee 1996.

Acknowledgements Bill Dunham, Florence Fasanelli, Stacy Langton, Jim Tattersall, Steven Weintraub, and two referees all deserve credit for the many suggestions they have made which have improved this paper.

References

Bullynck, Maarten (2010) Factor tables 1657–1817, with notes on the birth of number theory. Rev. Histoire Math., 16, 133–216.

Burton, David M (1976) Elementary number theory. Boston: Allyn and Bacon.

Calinger, Ronald S. (2016) Leonhard Euler: Mathematical Genius in the Enlightenment. Princeton.

Chabert, Jean-Luc (1999) A History of Algorithms. From the Pebble to the Microchip. Springer.

E26. Euler, L., Observationes de theoremate quodam Fermatiano aliisque ad numeros primos spectantibus (Observations on a theorem of Fermat and others on looking at prime numbers). Commentarii academiae scientiarum Petropolitanae, 6, 1738, 103–107. Reprinted Opera Omnia, Series 1, Volume 2, 1–5. There are two English translation on the Euler Archive, one by Jordan Bell, the other by David Zhao.

E134, Euler, L., Theoremata circa divisores numerorum (Theorems on divisors of numbers). Novi commentarii academiae scientiarum Petropolitanae, 1, 1750, 20–48. Reprinted Opera Omnia, Series 1, Volume 2, 62–85. English translation on the Euler Archive by David Zhao.

E228. Euler, L., De numeris, qui sunt aggregata duorum quadratorum (On numbers which are the sum of two squares). Novi commentarii academiae scientiarum Petropolitanae, 4, 1758, 3–40. Reprinted Opera Omnia, Series 1, Volume 2, 295–327. The original publication of E228 and an English translation by Paul Bialek are available on the eulerarchive.maa.org.

E241. Euler, L., Demonstratio theorematis Fermatiani omnem numerum primum formae $4n + 1$ esse summam duorum quadratorum (Proof of a theorem of Fermat that every number of the form $4n+1$ can be given as the sum of two squares), Novi commentarii academiae scientiarum Petropolitanae, 5, 1760, 3–13. Opera Omnia: Series, Series 1, Volume 2, 328–337

E283. Euler, L., De numeris primis valde magnis (On very large prime numbers). Novi commentarii academiae scientiarum Petropolitanae, 9, 1764, 99–153. Reprinted Opera Omnia, Series 1, Volume 3, 1–45.

E467. Euler, L., De tabula numerorum primorum usque ad millionem et ultra continuanda, in qua simul omnium numerorum non primorum minimi divisores exprimantur (On the table of prime numbers continued up to one million and beyond, in which at once all the non-prime numbers are expressed by their smallest divisors) Novi commentarii academiae scientiarum Petropolitanae, 19, 1775, 132–183 Reprinted Opera Omnia, Series 1, Volume 3, 359–404.

E699. Euler, L., Utrum hic numerus 1000009 sit primus necne inquiritur (An inquiry into whether or not 1000009 is a prime number). Nova acta academiae scientarum imperialis Petropolitinae, 10, 1797, 63–73. Reprinted Opera Omnia, Series 1, Volume 4, 245–254. An English translation by Jordan Bell is on the Euler Archive.

EG Corr. Leonhardi Euleri Commercium Epistolicum cum Christiano Goldbach. 2015. Opera Omnia IVA4, two volumes. Lemmermeyer, Franz and Mattmüller, Martin (editors). The letters are in Latin, German, and one in French, with English translation, and with abundant annotations by the editors.

Goldman, Jay R. (1998) The Queen of Mathematics. A Historically Motivated Guide to Number Theory, Wellesley, MA; A. K. Peters.

Heath, Thomas L. (1910) Diophantus of Alexandria. A Study in the History of Greek Algebra. The first edition was published by Cambridge University Press in 1885. We cite the second edition of 1910 as republished by Dover in 1964.

Jameson, G. J. O (2010) Two squares and four squares: the simplest proof of all? The Mathematical Gazette, vol. 94, no. 529, 119–123.

McKee, James (1996) Turning Euler's Factoring Method into a Factoring Algorithm. Bulletin of the London Mathematical Society 28 (volume 4); 351–355. An algorithm is presented which, given a positive integer n, will either factor n or prove it to be prime. The algorithm takes $O(n^{1/3} + \epsilon)$ steps.

Ore, Oystein (1948) Number Theory and Its History, McGraw-Hill. See pp. 59–64.

Riesel, Hans (1985) Prime Numbers and Computer Methods for Factorization. Birkhäuser. Discusses Euler's Factoring Method on pp. 158–159, but the techniques are somewhat different than Euler's.

Sandifer, C. Edward (2015) Factoring F_5, pp. 41–44 in How Euler Did Even More. MAA.

Weil, André (1984) Number Theory. An approach through history from Hammurabi to Legendre. Springer. J. J. Burckhardt has prepared Euler's work on number theory: A concordance for A. Weyl's Number Theory Historia Mathematica, 13 (1986), 28–35. This is very useful, for it links Eneström numbers (with page numbers in Euler's *Opera Omnia*) and page numbers in Weil.

Wagon, Stan (1990) The Euclidean algorithm strikes again. The American Mathematical Monthly, 97, 125–129.

Wertheim, Gustav (1887) Elemente der Zahlentheorie. Teubner. See pp. 295–299.

Zagier, D. (1990) A One-Sentence Proof That Every Prime $p \equiv 1 \pmod 4$ is a Sum of Two Squares. The American Mathematical Monthly, 97, 144.

“A Most Elegant Property”: On the Early History of Lexell’s Theorem

Eisso J. Atzema

Abstract In the late 18th century, both Anders Lexell (1740–1784) and his mentor Leonhard Euler (1707–1783) published a proof of a result in spherical geometry that is now known as Lexell’s Theorem. In this paper, I discuss both proofs and provide some background to the theorem. I will also outline some of the later proofs of the theorem as well as discuss some applications. In this connection, I argue that some of the work of Euler’s disciples Fuss and Schubert may be connected to Lexell’s investigations. In conclusion, I touch upon the role of Lexell’s Theorem in the history of hyperbolic geometry.

Keywords Spherical geometry • Spherical trigonometry • Lexell • Euler

1 Introduction

Sometime in 1777 or 1778, the Swedish astronomer and mathematician Anders Lexell (1740–1784) presented a paper to the Petersburg Academy of Sciences with the statement and a proof of what is now known as Lexell’s Theorem, a result regarding spherical triangles with a fixed base and a fixed area. Intrigued by Lexell’s findings, his famous colleague and close collaborator Leonhard Euler (1707–1783) tried his own proof, which was presented shortly afterwards and only published many years later. In this paper, I will discuss the proofs of both Lexell and Euler of the result published by the former. In conclusion, I will touch upon the work of Nicholas Fuss (1755–1826), another mathematician with close ties to Euler. In the early 1780s, Fuss also worked on spherical triangles and at least one of the questions he tackled seems to have been inspired by Lexell’s work. In fact, I will argue that the particular approach that Lexell followed to prove his eponymous theorem suggests

E.J. Atzema (✉)
University of Maine, Orono, ME 04469, USA
e-mail: eisso.atzema@maine.edu

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*, Proceedings of the Canadian Society for History and Philosophy of Mathematics La Société Canadienne d’Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_8

a much more direct proof of Fuss' result than the proof that Fuss himself published. The inspiration for this paper arose from Papadopoulos (2014), which has a broader scope, but covers much of the same grounds with regard to Lexell's Theorem. I feel, however, that my take on the theorem's history is sufficiently different from Papadopoulos' analysis for the present paper to be more than a mere footnote to his paper. Finally, I opted to redraw all of the diagrams discussed in this paper in GeoGebra, rather than to include the original illustrations. While a discussion of the lack of sophistication of Lexell's diagrams might be of some interest, I felt that their relative crudeness might be getting in the way of the story that I wanted to tell.

2 Anders Lexell

The Finnish-Swedish mathematician Anders Johan Lexell was born in Åbo (now Turku, Finland) in 1740.[1] After having completed his training in the sciences at his hometown university, he moved to Uppsala to become a mathematics lecturer in 1763. Later on he worked at the Uppsala nautical school. In 1768, at the suggestion of Euler, he applied for a position at the St. Petersburg Academy of Sciences. The following year, he was offered a position on the strength of his research and he would serve at the Academy in various capacities until his untimely death due to medical complications after an eye operation in 1784. For most of that time, he would work closely with Euler, who also became a personal friend. In fact, Lexell was one of the few witnesses to Euler's death and his account of the great mathematician's last hours was widely published. Most of Lexell's work was in astronomy, but he also published in pure mathematics, mostly in differential equations and geometry.

Lexell's papers on geometry cover a variety of topics, with a focus on the metrical properties of triangles and quadrilaterals. He had a special interest in spherical geometry, which may be defined here as the study of the geometrical properties of the surface of a sphere and any curves on it. Probably his best known result in this area (and the topic of this paper) is his discovery of the theorem now usually named after him in Lexell (1781).

3 Spherical Geometry Before Lexell

The notion of spherical geometry goes back to classical antiquity. Noting that great circles on a sphere have many properties in common with straight lines in the plane, astronomers such as Theodosius of Bithynia (160–100 BCE) and Menelaus of Alexandria (70–140) sought to formulate a framework for spherical geometry along

[1] On Lexell's life and work, see Stén (2014).

the lines of Euclid's *Elements*. Known as *spherics*, the resulting body of knowledge was considered part of astronomy and mostly served the needs of astronomers. After the demise of the Greek mathematical tradition, spherics was for the most part incorporated in the Muslim mathematical tradition, in which context it played an important role in the crystallization of trigonometry. By the 16th century, the study of spherical geometry had essentially transformed into spherical trigonometry and was mostly focused on techniques for "solving" spherical triangles. Certainly, some new properties in spherical geometry proper were still discovered, notably the formula for the area of a spherical triangle by Albert Girard and others. Mostly, however, these discoveries were made within the context of spherical trigonometry.

The Italian mathematician Vincenzo Viviani (1622–1703) broke new grounds in 1692, well before space curves were formally introduced as an object of study by Alexis Clairaut (1713–1765) in his *Recherches sur les courbes à double courbure* (1731), when he studied the properties of the curve obtained as the intersection of a cylinder with a sphere tangent to it where the radius of the latter is twice that of the former. Almost four decades later, in 1728, Viviani's student Guido Grandi (1671–1742) discussed various flower-shaped curves on a sphere in Grandi (1728). Around the same time, in the context of the study of gears, Jacob Hermann (1678–1733) introduced the so-called spherical epicycloid as the locus of a point on the edge of the base of a circular cone with its vertex in a fixed position as the cone is rolled along a plane in arbitrary position with respect to its vertex. In the 1730s, Johann Bernoulli (1667–1748) and Alexis Clairaut would contribute to this topic as well.[2]

A separate development involved the use of so-called *differential analogies* or expressions for the ratios of infinitesimal changes of various measurements concerning spherical triangles. This technique was pioneered by Roger Cotes (1682–1716) in his posthumously published *Harmonia Mensurarum* (1722), perfected by the French astronomer Nicolas-Louis de Lacaille (1713–1762), and put to good use by a whole host of other astronomers and mathematicians (including Euler).[3] Although the emphasis of this technique was still very much on the study of triangles, its exploitation of infinitesimal changes on a spherical surface may have helped to pave the way for the rise of differential geometry and the study of general surfaces by the late 18th century.

4 Lexell's Second Paper on Spherical Geometry

Most likely, Lexell's interest in spherical geometry was spurred by his close collaboration with Euler in St. Petersburg. As an astronomer, he clearly had to be well versed in spherical trigonometry. His first paper on spherical geometry,

[2] On Viviani, see Roero (1990). For further references, see Gregory (1815), p. 439 and Delcourt (2013).

[3] On this, see Delambre (1827) for a quick overview.

however, was a continuation of Hermann's work on spherical epicycloids and had little to do with spherical triangles. The connection to spherical trigonometry is clearer in his final two papers of the subject, as he seeks to extend the laws of spherical trigonometry to spherical polygons. His second paper, which is the main focus here, falls somewhere in the middle.[4]

We actually do not know when this second paper was written, although Euler's reworking of its contents suggests that it was presented before 1778.[5] At any rate, the paper was certainly presented and probably prepared for publication before Lexell's 18 months' long Grand Tour through Europe starting in July 1780. Due to the publication backlog of the St. Petersburg *Acta*, it was not actually published until shortly before Lexell's death in 1784.

In the introduction to the paper, Lexell notices that whereas spherical trigonometry has been extensively studied, the properties of curves on the sphere have not. If they were, however, surely this would lead to a new kind of geometry that would provide much of interest—even though not necessarily on the practical level. The purpose of his paper, he claims, is to provide an example of a problem that might be representative for this new kind of geometry.

Specifically, consider a fixed line segment AB in the plane and let C be any point in the plane such that the area of triangle ABC is of a given magnitude. Then, as shown already by Euclid in Propositions 37 & 39 of Book I of his *Elements*, the locus of C is a straight line parallel to the segment AB.[6] The sole topic of Lexell's paper is a discussion of the analogous question for spherical triangles. In other words, what is the locus of the vertices of all spherical triangles with a fixed base and a fixed area?

In order to solve this question, Lexell starts out with a spherical triangle with an arc AB for its base and a point V for its third (moveable) vertex, with C the midpoint of AB, R the foot of the altitude from V, Z the pole of the great circle of AB on the side of V, and M the antipode of C. Furthermore, he denotes the lengths of the arcs CR, RV, and VC by x, y, and z, respectively. Finally, μ refers to the angle $\angle AVC$, ν to $\angle BVR$, and ϕ to $\angle VCZ$ (see Figure 1).

Lexell's key tool in answering his own question is now one of Napier's four so-called Analogies,[7] as first given by John Napier in 1614 and supplemented by Henry Briggs in 1617. This particular result states that for any spherical triangle with angles A, B, C and opposite sides a, b, c, one has the equality

$$\frac{\tan(\frac{1}{2}A + \frac{1}{2}B)}{\cot(\frac{1}{2}C)} = \frac{\cos(\frac{1}{2}a - \frac{1}{2}b)}{\cos(\frac{1}{2}a + \frac{1}{2}b)}.$$

4 See Lexell (1779), Lexell (1781), Lexell (1782a), and Lexell (1782b).

5 On this, see Footnote 10.

6 Strictly speaking, unless we want to bring in the notion of oriented area, the locus consists of two lines on opposite sides of AB as C can lie on either side of AB. Euclid avoids this issue by only talking about triangles on the same side of the fixed base.

7 From the Greek "ana" (again) and "logos" (ratio), i.e., equalities between ratios.

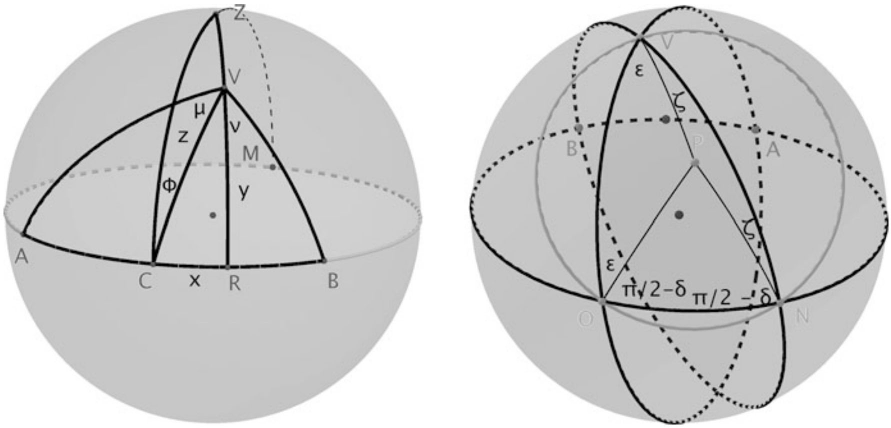


Fig. 1 Lexell's Trigonometrical Solution and Geometrical Construction

By applying this relation to ARV and BRV, Lexell obtains the relations

$$\tan(\frac{1}{2}A + \frac{1}{2}\mu) = \frac{\cos(\frac{1}{2}y - \frac{1}{2}a - \frac{1}{2}x)}{\cos(\frac{1}{2}y + \frac{1}{2}a + \frac{1}{2}x)},\ \tan(\frac{1}{2}B + \frac{1}{2}\nu) = \frac{\cos(\frac{1}{2}y - \frac{1}{2}a + \frac{1}{2}x)}{\cos(\frac{1}{2}y + \frac{1}{2}a - \frac{1}{2}x)},$$

where a is the length of the arc AB. After some tedious manipulation, these equations lead to the relation

$$\cot(\delta)\sin(a)\sin(y) = \cos(y)\cos(x) + \cos(a),$$

where 2δ is the so-called spherical excess $A + B + V - \pi$ of ABV, which also equals the area of ABV by Girard's Theorem. As $\cos(y)\cos(x) = \cos(z)$ and $\sin(z)\cos(\phi) = \sin(y)$, he now has

$$\cot(\delta)\sin(a)\sin(z)\cos(\phi) = \cos(z) + \cos(a).$$

Setting $\tan(\epsilon) = -\cot(\delta)\sin(a)$ and $\cos(\gamma) = -\cos(a)\cos(\epsilon)$, this is equivalent to the relation

$$\cos(\gamma) = \cos(z)\cos(\epsilon) + \sin(z)\sin(\epsilon)\cos(\phi).$$

Note that both γ and ϵ only depend on δ and a. In other words, they are the same for any V such that ABV has area 2δ. Now, let P be the (unique) point on CZ such that the arc CP has length ϵ. Then, as the right-hand side of this equation can be interpreted as an expression for the cosine of the arc VP by the law of cosines applied to triangle CVP, it follows that the length of the arc VP is the same for all V such that AVB has a fixed area. Specifically, any V such that ABV has area 2δ has to lie on a circle with center P. This proves part of what we now know as Lexell's Theorem.

Now that he has shown that the locus of V is a circle, Lexell's next step is to construct the circle that he found as the locus for all V such that the area of AVB is fixed. As before, assume that the area of ABV has to equal 2δ and let O and N be the antipodes of A and B, respectively. Now erect an isosceles triangle on ON such that the base angles equal $\frac{1}{2}\pi - \delta$, then, Lexell claims, the third vertex of this triangle is the center P of the circular locus of V and the circle itself passes through O and N. This is the full statement of what is now known as Lexell's Theorem. Actually, it is a little bit more than the usual statement, as the part about the base angles of OPN is generally not included.

Lexell gives two proofs of his statement, one mostly trigonometric and the other purely geometric. We will omit the former here and only discuss the latter. Compared to his trigonometric proof, this geometrical argument is surprisingly short and has quite a classical (Euclidean) ring to it—even if Lexell's write-up could have used some streamlining. Essentially, his line of reasoning is as follows. Let V be any point on the circle with center P (constructed as the top of an isosceles triangle APB) and passing through O and N such that P lies inside triangle ONV (see Figure 1). Now, let ϵ denote the magnitude of the base angles of the isosceles triangle OPV, while ζ denotes the magnitude of the base angles of the isosceles triangle NPV. Then, by construction, $\angle BAV$ is supplementary to $\angle NOV$, while $\angle ABV$ is supplementary to $\angle ONV$. In contrast, $\angle BVA$ is congruent to $\angle NVO$. Therefore, the spherical excess of ABV equals $\pi - (\epsilon + \frac{1}{2}\pi - \delta) + \pi - (\zeta + \frac{1}{2}\pi - \delta) + \epsilon + \zeta - \pi = 2\delta$. In other words, for any such V as above, the area of triangle ABV equals 2δ, which shows that V has the desired property. When V is such that P does not lie inside triangle OPV (with V still on the same side of the great circle AB as P), a slight variation of the argument just given shows that in that case V has the desired property as well.

Although Lexell's geometrical proof is entirely original and does not seem to be modeled after anything Euclid does, it does parallel Proposition 37 of Book I of his *Elements* ("Triangles which are on the same base and in the same parallels are equal to one another"). Just as Euclid's proposition, it does not prove that the locus of V is just the circle with center P and passing through O and N, so it does not completely replace his trigonometric line of reasoning. On the other hand, a proof paralleling the contrapositive argument of Euclid's Proposition 39 of Book I of the *Elements* ("Equal triangles which are on the same base and on the same side are also in the same parallels.") certainly could have.[8] Thus, in effect, Lexell provided a purely geometrical proof of the theorem now called after him.

[8]Indeed, let V' be any point not on the circle with center P and passing through O and N, but on the same side of AB as P. Then, there is a unique circle through V', O, and N with its center P' on the perpendicular bisector of ON. If the area of ABV' equals $2\delta'$, it follows that $\angle P'ON$ equals $\frac{1}{2}\pi - \delta'$. Since P and P' will be distinct, so will δ and δ'. In other words, V' does not have the desired property.

5 Leonhard Euler's Contribution

The belated publication of Lexell's paper does not seem to have attracted much attention. In fact, the only known immediate reaction to the paper came before it was even published—probably in response to its presentation to the St. Petersburg Academy. We already mentioned that Lexell's work may have been inspired by Euler's interest in spherical geometry. Specifically, Euler was interested in the foundations of spherical trigonometry, a topic that the Swiss mathematician had turned to again in the mid-1770s and would work on on-and-off until his death in 1784. At this point, given that he was blind and increasingly deaf, Euler no longer was attending the meetings of the Academy,[9] but surely he knew Lexell's work first-hand through their close collaboration. One of Euler's papers on the topic of this period is Euler (1797), a long paper on the use of differential analogies to find expressions for the area of a spherical triangle and presented to the Academy on January 29, 1778. Near the end of this paper, Euler points out that his deliberations on differential analogies can also be used to derive Lexell's main result.[10] Claiming that his derivation of Lexell's "most elegant property" (*elegantissima proprietas*) cannot be done easily, he never actually shows how to do this. Instead, he gives an alternative purely geometric derivation.

Unlike Lexell's proof, Euler's argument very closely models Euclid's proof of Proposition 37. His first observation is that for any spherical triangle EFe, we can draw the great circle through the midpoints of the sides Ee and Fe and then think of E and F as positioned on a small circle parallel to the great circle and e as a point on the small circle that is the reflection of the circle of E and F in the plane of the great circle (see Figure 2). Now let f be the reflection of E in the midpoint of Fe. Then, by construction, f lies on the same circle as e and the area of the figure $EFfeG$ (bounded by two great circles and two small circles, which Euler refers to as a "spherical parallelogram") is twice the area of EFe. Now let ϵ be another point on the circle of e and f and let ζ be the reflection of E in the midpoint of $F\epsilon$. Then, the area of the "spherical parallelogram" $EF\zeta\epsilon$ is twice that of $EF\epsilon$. Now, comparing angles, the "triangles" $E\epsilon e$ and $F\zeta f$ (each bounded by two great circles and one small circle) are congruent and therefore of the same area. But then, as $EF\zeta\epsilon = E\epsilon e + EFO - O\epsilon\zeta$ and $EFfe = F\zeta f + EFO - O\epsilon\zeta$, the areas of the "parallelograms" $EFfe$ and $EF\zeta\epsilon$ and therefore the areas of the "triangles" EFe and $EF\epsilon$ are equal as well. Consequently, if we replace the small circle from E to F with the great circle determined by E and F, the spherical triangles EFe and $EF\epsilon$ have the same area as well. As the antipodes of E and F also lie on the small circle of e and ϵ, it follows that all points ϵ on the small circle through e and the antipodes of E and F are such that the triangle $EF\epsilon$

[9] See Calinger (2016), p. 515.

[10] Euler does not actually indicate how he knew about Lexell's result. He just refers to it as a theorem "proposed" by Lexell. If this is a reference to the presentation of Lexell's paper, it would place the presentation before January 1778 (assuming that the pertinent part of Euler's paper is not a later insertion).

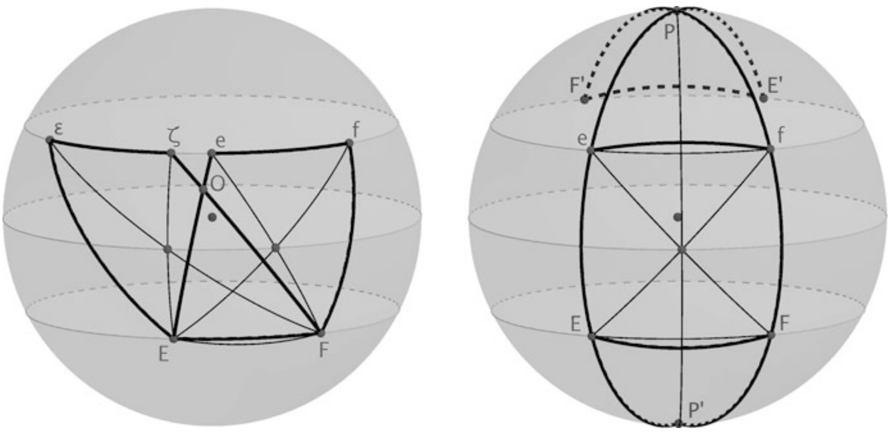


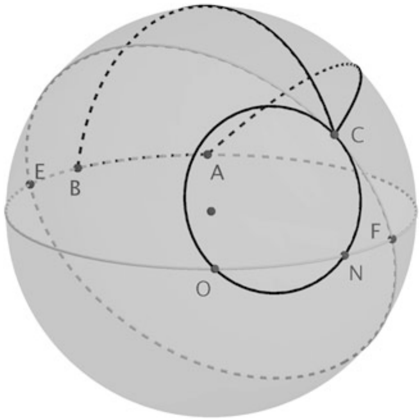
Fig. 2 Euler's Geometrical Proof

has the same area as *EFe*. Finally, he shows that this small circle is the same small circle that Lexell found. In order to do so, he defines P' as the center of the small circle through E and F, with P its antipode and picks e on the great circle through E and P and therefore f on the great circle through F and P, making the triangles *EPF* and *ePf* isosceles. For $\angle EPF$ measured in radians, it follows that the lune $PEP'F$ has area $2\angle EPF$. By construction, however, that area also equals twice the sum of the area of *EFe* (i.e., 2δ) and the area of *efP* (i.e., $\angle EPF + 2\angle feP - \pi$, by Girard's Theorem). Therefore, $\angle feP = \frac{1}{2}\pi - \delta$, which defines the circle and which is the same relation that Lexell found for his circle. This completes the proof of Lexell's Theorem. Note that along the way, Euler shows that for all triangles *EFe* with fixed base *EF* and fixed area, the locus of the midpoints of *Ee* is a straight line coinciding with the locus of the midpoints of *Fe*—even though it is not clear whether he noted this himself. Just like Lexell in his geometrical proof, Euler actually never shows that this small circle is the complete locus.

6 A Paper by Nicholas Fuss

Euler's paper only appeared in print in 1797. In the meantime, no other papers related to Lexell's discovery were published with the possible exception of a paper by the Swiss-Russian mathematician Nicholas Fuss. Born, raised and academically trained in Basel, Fuss had come to Saint Petersburg in 1773 as a personal secretary to the aging Euler. For the last decade of Euler's life, Fuss was responsible for the write-up and presentation to the Academy of most of Euler's papers. After Euler died, Fuss stayed on and for the last 25 years of his life he was the permanent secretary of the St. Petersburg Academy. In 1786, roughly two years after both Lexell and Euler had died, Fuss published Fuss (1786) in which he gives

Fig. 3 Fuss' Problem



a construction for the unique spherical triangle ABC with given base AB and C on a given great circle, such that ABC is of maximal area. Certainly this kind of problem seems to fall in the same category as Lexell's and Euler's investigations. Indeed, given that Euler, Lexell and Fuss had closely collaborated on other problems, it seems likely that the three of them would have discussed their work in spherical geometry as well. The actual paper, however, is not geometrical in style at all and uses spherical trigonometry only. Intriguingly, though, Lexell's geometrical construction of the small circle called after him provides an almost immediate solution to the question that Fuss addresses.

In fact, let E be the point of intersection of the given line and the line of AB that lies closest to AB (see Figure 3). Let F be the antipode of E, while (as before), O and N are the antipodes of A and B. By Lexell's Theorem, for all triangles ABV with fixed base AB and a given area have, the moveable vertex V lies on the same circle through O and N. In general, therefore, there will be two points V on the given line such that ABV has a given area. By Lexell's construction, the bigger the given area of the triangle with base AB, the smaller the radius for the corresponding circle through O and N. Therefore, these two points on the given line should merge for the triangle with the biggest possible area. In other words, the vertex C of the triangle of maximal area should be the point of tangency of the unique circle through O and N touching the given line. By the power property of spherical circles, it now follows that $\tan^2(\frac{1}{2}FC) = \tan(\frac{1}{2}FO) \cdot \tan(\frac{1}{2}FN)$ or $\cot^2(\frac{1}{2}EC) = \tan(\frac{1}{2}EA) \cdot \tan(\frac{1}{2}EB)$. Fuss has the same result, but does not follow the same approach, nor could he really have as the notion of the power of a spherical circle had not been introduced yet.[11] This suggests that, perhaps, Fuss did try to use Lexell's construction, but reached an impasse with the actual construction of C. His paper, then, could be an attempt to solve the problem by other means. At

[11] According to Max Simon, the power property of a spherical circle was first expressed in Cagnoli (1804), p. 349. See Simon (1906), p. 247.

the very least, the problem that he tackled seems to have been inspired by Lexell's work. In Fuss (1784), an expanded version of his paper, presented on June 11 of 1786, Fuss discusses two more similar problems, but neither seems to lend itself to a short geometrical characterization and Fuss' findings strongly suggest there might not be an elegant solution at all. Incidentally, in Schubert (1786), presented to the St. Petersburg 11 days after Fuss (1784), Fuss' sometime collaborator (and maternal great-grandfather of Sofia Kovalevskaya) Friedrich Theodor Schubert (1728–1825) addressed a variation of Fuss' question with the alternate restriction that the vertex C be on a small circle parallel to the great circle of AB. In this case, it readily follows from Lexell's construction that the maximum area is achieved for C at the point of intersection of the perpendicular bisector of AB with the small circle that lies furthest away from AB, while the minimum is achieved for the other point of intersection. Just as Fuss, however, Schubert follows a more workman-like, calculus-based approach, ultimately arriving at the same answer.

7 19th-Century Reception

Just like the published version of Lexell's paper, the belated publication of Euler's paper does not seem to have had much of an immediate impact. Two years after Euler's paper was published, the French mathematician Adrien-Marie Legendre (1752–1833) decided to include Lexell's Theorem in the third edition of his influential geometry textbook (see Legendre (1800), pp. 358–9). The theorem, however, was not incorporated in the main text. Instead it was tagged on to an existing note on the surface area of a spherical triangle—one of several notes included as appendices to the actual text. Interestingly, Euler is not mentioned at all and the reference to Lexell's paper refers to the 5th volume of *Nova Acta* (of 1787) instead of the *Acta* (of 1782). There is hardly any question, however, that Legendre read Lexell's work. His strictly trigonometric proof is clearly modeled after Lexell's trigonometric proof—with some clever use of a formula for the area of a spherical triangle derived earlier in the note.

Although Legendre's textbook was highly influential, the strictly trigonometric nature of his proof of Lexell's Theorem and the fact that it was hidden at the end of one of the notes probably did not help to attract attention to Lexell's work. Indeed, it would take almost 30 years before any other discussion of Lexell's circle made it into print. In 1827, in Steiner (1827), mostly a paper on the dissection of spherical figures, the Swiss-German mathematician Jacob Steiner (1796–1863) showed how Lexell's Theorem could be put to use. Somewhat dismissive of Legendre's proof and clearly unaware of Lexell's and Euler's proofs, he also provided his own (rather short) geometrical proof of the theorem. The key to his proof is the observation that for any spherical quadrilateral, the sum of the angle of the one pair of opposite angles has to equal the sum of the other pair of angles. Now, consider a triangle $A'B'C$ with a point D on the circumcircle of $A'B'C$ (with the tacit assumption that A' and B' separate C and D) (see Figure 4). As usual, let the angles of $A'B'C$ be denoted

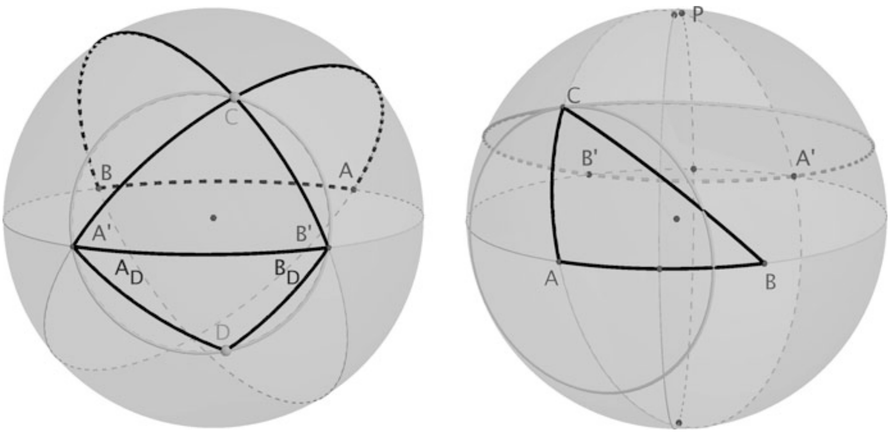


Fig. 4 Steiner on Lexell’s Circle

by the labels of the corresponding vertices, while the angles in $A'B'D$ are denoted by A_D, B_D and D. Then, $A'+A_D+B+B_D = C'+D$ or $A'+B'-C = D-A_D-B_D$. Now, this same relation will hold if we replace C with any other point on the circumcircle of $A'B'C$ (again assuming that this point is separated from D by A' and B'). As the right-hand side of this expression does not change, it follows that the locus of all triangles $A'B'C$ with a fixed basis $A'B'$ with a fixed value for the expression $A' + B' - C$ is a circle passing through A' and B'. Now, let A and B be the antipodes of A' and B' and let the angles of ABC be denoted in the usual way. The locus of all points C such that $A' + B' - C$ is constant now is a circle through A' and B'. But, as $A + A' = B + B' = \pi$, it follows that for each such C, the sum $A + B + C$ of the angles of ABC is constant. This proves Lexell’s Theorem.

Steiner never constructs Lexell’s circle in the way Lexell and Euler did. The only property of the circle that he needs is that the great circle tangent to Lexell’s circle at A' and the great circle of AC define a lune which has the same area as ABC and likewise for the tangent circle at B'. Using this observation, he now can construct for any spherical triangle ABC a line through each of the vertices that divides the triangle into two triangles of equal area and show that these three “equalizers” are actually concurrent. He also outlines several other dissection constructions based on this observation. Finally, after deriving some properties of spherical conic sections, he uses these results to solve an optimization problem not unlike Fuss’ problem.

In this case, the question is how to characterize the triangle of maximal area when two of its sides are given in length. It is not clear whether this problem was ever considered by Euler and his circle. The only reference Steiner gives is to Legendre’s *Élements de géométrie*, which has a rather long proof showing that for any such triangle the angle between the two given sides has to equal the sum of the other two angles.[12] A more succinct proof, however, follows almost directly from Lexell's Theorem by taking one of the given sides as the fixed base *AB*. If the other side of fixed length meets *AB* at *A*, the point *C* has to lie on a circle with center *A*. But then, the area of *ABC* will be maximal exactly when the corresponding Lexell circle touches this fixed circle with center *A*.[13] Consequently, for this triangle, the center *P* of its Lexell circle lies on the great circle of *AC* and the angle *A* of *ABC* is supplementary to the angle A' of $A'PB'$. By Lexell's construction, this angle also equals $\pi - \frac{1}{2}(A+B+C)$. In other words, $A = B+C$. Although Steiner's derivation of this result only uses Lexell's circle indirectly, the construction and proof he gives at the end is essentially the same as just sketched. Following Legendre, Steiner points out that this property implies that the center of the circumscribed circle of such a maximal circle has to coincide with the midpoint of the variable side.[14]

Although a seminal paper in other respects,[15] the publication of Steiner (1827) does not seem to have led to an increased awareness of Lexell's Theorem, although a short history of the theorem and Steiner's proof of it are discussed in Klügel et al. (1831) (pp. 250–252). Steiner's incorporation of much of the same material in Steiner (1841) of 1841—a winning submission to a prize contest of the Parisian Academy of Sciences, which incorporated much of Steiner (1827)—seems to have gone largely unnoticed as well.[16] In fact, on December 29 of 1841, Carl Friedrich Gauss (1777–1855) inquired of his correspondent Heinrich Christian Schumacher (1780–1850) whether it was a known result that the locus of the moveable vertex of all spherical triangles with a fixed base and area is a small circle. Exceedingly well-informed as always, Schumacher promptly replied on January 3 of 1842 with a reference to Steiner (1827), Legendre (1800), and a correct reference to Lexell (1781), while also sending along a separate (unprinted) proof by his then assistant

[12]The theorem was already included in the first edition of the book, but its very short proof was based on the unverified assumption that the maximum area is reached for exactly one configuration of the two given sides. Starting with the third edition, a much longer proof replaces the original (but incomplete) proof. See Legendre (1800), pp. 254–257.

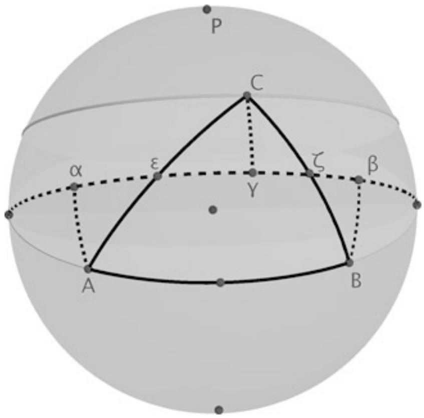
[13]Actually, this construction is only possible as long as the antipode B' of *B* lies "outside" the circle with center *A*, i.e., as long as the sum of the given sides is less than two right angles. If the sum of the given sides is greater or equal to two right angles, the area of the triangle increases as the angle between the given line segments increases until the triangle degenerates into three arcs on the same great circle. Steiner does not distinguish between these two cases, although Legendre does.

[14]Note that the latter characterization implies in its turn that the chordal (rectilinear) triangle *ABC* has to be right-angled at *A*. This would allow for an explicit construction of the spherical triangle *ABC* of maximal area, but clearly it is not a property intrinsic to spherical geometry as the study of the geometry of the *surface* of a sphere. In terms of spherical trigonometry, given that *A* has to equal $B + C$, it follows from the Napier analogy used by Lexell (see above) that $\cos(A) = -\tan(\frac{1}{2}b)\tan(\frac{1}{2}c)$, where *b* and *c* are the given lengths of the triangle. Since $b + c < \pi$, *A* is well-defined.

[15]Among other things, it established that any spherical figure bounded by great circles can be decomposed into any other such spherical figures of the same area. The corresponding result for the plane, also known as the Bolyai-Gerwien Dissection Theorem, was proved independently by Wolfgang Bolyai and Paul Gerwien in the 1830s, although the British mathematician William Wallace is said to have proved the same result in 1807.

[16]One exception is Catalan (1843), pp. 272–273, which follows Steiner's argument very closely.

Fig. 5 Gauss' Proof



Thomas Clausen (1801–1885).[17] As is clear from his response 3 days later, Gauss had his own proof as well. Although he does not give all the details, the basic idea is clear.

Starting with two points A and B on a great circle, Gauss considers a second great circle intersecting the circle of AB at two points equally far removed from the midpoint of AB (see Figure 5). By construction, the perpendiculars $A\alpha$ and $B\beta$ from A and B to this second circle will be congruent. Now, pick any point ϵ on the second great circle and extend $A\epsilon$ to a point C such that ϵ is the midpoint of AC. Again, by construction, the perpendicular $C\gamma$ from C to the second great circle will be congruent with $A\alpha$ and therefore with $B\beta$. Now, let ζ be the point of intersection of CB with the second great circle. Then, the triangles $C\gamma\zeta$ and $B\beta\zeta$ will be congruent, which implies that ζ is the midpoint of BC. Finally, assume that γ lies between ϵ and ζ. As $C\gamma\epsilon$ is congruent with $A\alpha\epsilon$ and $C\gamma\zeta$ is congruent with $B\beta\zeta$, it follows that the triangle ABC has the same area as the quadrilateral $AB\beta\alpha$. By a similar argument, it can be shown that the same is true when γ lies outside of $\epsilon\zeta$. As the area of $AB\beta\alpha$ is independent of our choice of ϵ, it follows that all triangles constructed in this way (with AB and the second great circle fixed) have the same area. But then, as Gauss points out, the locus of the vertices C of all such triangles will be a small circle equidistant to the great circle of α and β which has the pole of $\alpha\beta$ on the opposite side of A and B for its center. This completes Gauss' proof of Lexell's Theorem. From his reaction to Clausen's proof, it would appear that the latter was along similar lines.

Thus, almost 60 years after Lexell's original discovery of his circle, his "most elegant theorem" was known to some, but hardly part of mathematical lore. His original proof was first replaced with Euler's, then with Steiner's and finally with

[17]See Peters (1862), pp. 46–49. Schumacher's correction of the reference in Legendre (1800) to Lexell (1781) could be an indication that he also saw Steiner (1841), which has the correction in a footnote.

Gauss'—each extending Lexell's original statement ever so slightly. With Gauss' contribution, most aspects of Lexell's circle seem to have been explored. Other proofs would follow, but none seem to have added any fundamentally new insight, and modern proofs are essentially all variations of the approaches outlined in this paper.[18] In this sense, the evolution of Lexell's work was complete by the middle of the 19th century (even though Gauss' proof was to remain unpublished until 1862).

8 In Conclusion

When Lexell presented his circle as an example of a new geometry on the sphere, he seemed unsure about the relation of this new geometry to Euclid's geometry of the plane. It is not entirely clear whether the close parallels between his geometrical proof and Euclid's proof of Proposition 37 of Book 1 had more than a heuristic significance for Lexell. Likewise, even though Euler was able to prove Lexell's Theorem by mostly adapting Euclid's proof of Proposition 37 to the case of a spherical triangle, he does not seem to attach much significance to the similarities between his proof and Euclid's reasoning. Similarly, Lexell does not give any meaningful application for his newly found circle and neither does Euler. As I argue in this paper, the publications of Fuss and Schubert on certain optimization problems on a sphere may have been inspired by Lexell's Theorem, but we do not know that to be actually the case.

Steiner was really the first to find any use for Lexell's Theorem in his paper Steiner (1827) on spherical dissections, thus making the theorem an essential part of a purely geometric study of the sphere. At the same time, Steiner also managed to abstract the proof of Lexell's Theorem to the point that the actual embedding of the sphere in space hardly played a role anymore. By doing so, he put the spotlight on the close analogy between Lexell's Theorem and Euclid's Proposition 37 of Book 1, thus paving the way for a completely axiomatic proof of Lexell's Theorem.

Among the first to follow this axiomatic approach was Gauss—even though he seems to have found Lexell's Theorem independently of both Lexell and Steiner. Although we do not know why Gauss was interested in Lexell's locus, his approach provided the final step toward a determination of this locus not just from the axioms of spherical geometry (in which case the locus is a circle), but also from the axioms of hyperbolic geometry (in which case the locus is a hypercycle). In fact, as we now know, the exact same step was taken by Nikolai Lobachevsky (1792–1856) both to prove the theorem in spherical geometry and then to generalize it to his "imaginary" geometry in the 1830s, while working on his mathematics in relative obscurity.[19]

[18] See, for instance, Le Besgue (1855), Barbier (1864), Persson (2012), and Maehara and Martini (2017) (which appeared after this paper had been submitted).

[19] See Engel and Stäckel (1898), p. 321. It seems unlikely that at the time Gauss was aware of these proofs. On the other hand, a summary of Lobachevsky's work on non-Euclidean geometry had

Thus, more than half a century after its discovery, Lexell's Theorem had found new purpose within the growing field of non-Euclidean geometry–even though its standing remained somewhat marginal. In fact, although Gauss' diagram kept popping up in the literature, it rarely was in the context of Gauss' original question.[20] Yet, in many instances of the use of Gauss' diagram, Lexell's Theorem could have been added as a straightforward corollary of the result that was proved. While one could argue that the axiomatic formulation of non-Euclidean geometry in the early 19th century to some extent was heralded by Lexell's call for a new kind of geometry of the sphere, the priorities set within this new field of study also reduced the theorem that he had offered as an illustration of what he had in mind to barely more than a footnote.

References

Barbier, J.É.: Démonstration du théorème de Lexell (Nova Acta Petropolitanae, vol. v). Les Mondes **4**, 42–43 (1864)

Cagnoli, A.: Trigonometria piana e sferica. Edizione seconda. Notabilemente ampliata. I Fratelli Masie Comp., Bologna (1804)

Calinger, R.S.: Leonhard Euler. Mathematical Genius in the Enlightenment. Princeton University Press (2016)

Catalan, E.: Éléments de géométrie. Bachelier, Paris (1843)

Delambre, J.B.J.: Histoire de l'astronomie au dix-huitième siècle. Bachelier, Paris (1827)

Delcourt, J.: Les epicycloides sphèriques. http://clea-astro.eu/conferences/fichiers/clairaut/jean-delcourt (2013)

Engel, F., Stäckel, P. (eds.): Urkunden zur Geschichte der Nichteuclidischen Geometrie I. Nikolaj Iwanowitsch Lobatschefskij. B. G. Teubner, Leipzig (1898)

Euler, L.: Variae speculationes super area triangulorum sphaericorum. Nova acta academiae scientiarum imperialis petropolitanae **10**, 47–62 (1797)

Fuss, N.: Problematum quorundam sphaericorum solutio. Nova acta academiae scientiarum imperialis petropolitanae **2**, 67–80 (1784)

Fuss, N.: Leichte Auflösung eines Problems aus der höheren Sphärischen Trigonometrie. Leipziger Magasin für reine und angewandte Mathematik **2**, 241–245 (1786)

Grandi, G.: Flores geometrici ex rhodonearum, et cloeliarum curvarum descriptione resultantes. Tartini & Franchi, Florence (1728)

Gregory, O.: A Treatise on Mechanics, Theoretical, Practical, and Descriptive, vol. 2, 3rd edn. Rivington (1815)

Hilbert, D.: Foundations of Geometry. Open Court, La Salle (1971)

---

been published in German in 1840. Even if the exact chronology is not entirely clear, it is tempting to think that Gauss' request was inspired by this summary and his own interest in non-Euclidean geometry.

[20] In 1844, very similar diagrams are used a number of times in Meikle (1844) to argue that the sum of the angles of a triangle is a measure for the area of that triangle. Likewise, Hilbert uses the same diagram in his proof of Legendre's Second Theorem ("If in some triangle the sum of its angles is equal to two right angles, then the same is true in every triangle") in the later editions of his *Grundlagen der Geometrie* (see Hilbert (1971), p. 38).

Klügel, G.S., Mollweide, C.B., Grunert, J.A.: Mathematisches Wörterbuch oder Erklärung der Begriffe, Lehrsätze, Aufgaben und Methoden der Mathematik den nöthigen Beweisen und literarischen Nachrichten begleitet in alphabetischer Ordnung. Erste Abtheilung. Die reine Mathematik. Fünfter Theil, vol. 1. E. B. Schwickert (1831)

Le Besgue, V.A.: Démonstration du théorème de Lexell. Nouvelles Annales des Mathématiques **14**, 24–26 (1855)

Legendre, A.M.: Éléments de géométrie. Avec des notes., 3rd edn. Firmin-Didot, Paris (1800)

Lexell, A.: De epicycloidibus in superficie sphaerica descriptis. Acta academiae scientiarum imperialis petropolitanae **3**(1), 49–71 (1779)

Lexell, A.: Solutio problematis geometrici ex doctrina sphaericorum. Acta academiae scientiarum imperialis petropolitanae **5**(1), 112–126 (1781)

Lexell, A.: De proprietatibus circulorum in superficie sphaerica descriptorum. Acta academiae scientiarum imperialis petropolitanae **6**(1), 157–190 (1782)

Lexell, A.: Demonstratio nonnullorum theorematum ex doctrina sphaerica. Acta academiae scientiarum imperialis petropolitanae **6**(2), 85–95 (1782)

Maehara, H., Martini, H.: On Lexell's Theorem. American Mathematical Monthly **124**(4), 337–344 (2017)

Meikle, H.: On the theory of parallel lines. The Edinburgh New Philosophical Journal **36**, 310–318 (1844)

Papadopoulos, A.: On the works of Euler and his followers on spherical geometry. Gaṇita Bhārātī **36**(1), 53–108 (2014)

Persson, U.: Lexell's theorem. Normat **60**(3), 133–134, 144 (2012)

Peters, C.A.F. (ed.): Briefwechsel zwischen C. F. Gauss und H. C. Schumacher, vol. 4. Gustav Esch (1862)

Roero, C.S.: Leibniz and the temple of Viviani : Leibniz's prompt reply to the challenge and the repercussions in the field of mathematics. Annals of Science **47**(5), 423–443 (1990)

Schubert, F.T.: Problematis cuiusdam sphaerici solutio. Nova acta academiae scientiarum imperialis petropolitanae (pars historiae) **4**, 89–94 (1786)

Simon, M.: Über die Entwicklung der Elementar-Geometrie im XIX. Jahrhundert. B. G. Teubner, Leipzig (1906)

Steiner, J.: Verwandlung und Theilung sphärischer Figuren durch Construction. Journal für die reine und angewandte Mathematik **2**, 45–63 (1827)

Steiner, J.: Sur le maximum et minimum des figures dans le plan, sur la sphère et dans l'espace général. Journal des mathématiques pures et appliquées **6**, 105–170 (1841)

Stén, J.C.E.: A Comet of the Enlightenment. Anders Johann Lexell's Life and Discoveries, *Vita Mathematica*, vol. 17. Birkhäuser (2014)

The Misnamings of Playfair's Axiom

Amy Ackerberg-Hastings

Abstract The term "Playfair's Axiom" is a mainstay of school geometry textbooks as well as one of the few things many mathematicians know about John Playfair (1748–1819), Professor of Mathematics and then of Natural Philosophy at the University of Edinburgh. However, the ubiquity of the phrase masks considerable historical complexity. At least three different versions of the statement circulate among speakers of English—only two of which appeared in the editions of Playfair's *Elements of Geometry*—while the underlying concept dates back to Proclus. Although historians and mathematicians have typically assumed that Playfair crafted axioms in response to 18th-century concerns with the problematic nature of the parallel postulate, recent research has revealed that "Playfair's Axiom" gained its name from English educators by 1830 and that the label was transferred to the third phrasing in the 1870s. Untangling the story of Playfair's Axiom thus enriches our understanding of the history of 19th-century mathematics education.

Keywords John Playfair • Euclidean geometry • Parallel postulate • Mathematics education

1 Introduction

As one of the few people in the world who study John Playfair (1748–1819), I receive a lot of questions about Playfair's Axiom—even if the topic of the talk I have just given had nothing to do with geometry. The interest in this statement is understandable. Searches in databases such as JSTOR, Google Books, and Google Scholar reveal at least 200 20th-century references in school and college geometry textbooks, articles on strategies for teaching geometry, and histories and popularizations of non-Euclidean geometry. Among mentions in the 21st century are those by Brooklyn College's Suzuki (2002, p. 648), Mazur, retired from Marlboro

A. Ackerberg-Hastings (✉)
Independent Scholar, Adelphi, MD
e-mail: aackerbe@verizon.net

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*, Proceedings of the Canadian Society for History and Philosophy of Mathematics La Société Canadienne d'Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_9

College (2006, pp. 76, 83, 287), and California State (Long Beach) professor Stein (2008, pp. 103–106). In each of these cases, with only minor differences in wording, the authors associate the label "Playfair's Axiom" with this statement:

> Given a line and a point not on the line, one and only one line can be drawn through the point that is parallel to the given line. (3)

For reasons given below, I will call this statement (3). If we check what John Playfair, then Joint Professor of Mathematics at the University of Edinburgh, actually wrote in the 1795 first edition of his influential textbook, *Elements of Geometry*, we find the sentence I will number statement (1):

> Two straight lines cannot be drawn through the same point, parallel to the same straight line, without coinciding with one another (Playfair 1795, p. 7). (1)

He revised his axiom about parallels in the 1804 second edition of *Elements of Geometry*, henceforth described as statement (2), but this sentence is also not the one that so many of us think we know (including from my experience as a high school sophomore in 1985–1986 studying a geometry textbook copyrighted in the 1950s):

> Two straight lines, which intersect one another, cannot be both parallel to the same straight line (Playfair 1804, p. 7). (2)

In other words, the statement which speakers of English typically associate with the name "Playfair's Axiom" today never appeared in Playfair's textbook. This oddity made it difficult to answer those questions at my conference presentations. David Henderson and Daina Taimina had also noticed the discrepancy between more recent textbooks and Playfair's. They further pointed out that the phrasing which is now widespread is not logically equivalent to either of the statements used by Playfair (Henderson and Taimina 2005, pp. 60–62). Statement (3) asserts both that parallel lines exist and that each pair is unique, while the axioms in *Elements of Geometry*, like Euclid's parallel postulate, posit only the uniqueness of parallel lines. (As they note, Euclid's proposition I.31 was designed to prove the existence of parallel lines.) Henderson and Taimina were particular about the logic claims because an axiom or postulate which posits only uniqueness is valid in non-Euclidean as well as Euclidean geometries, so existence must also be asserted for a statement to be true only in Euclidean geometry. When, how, and why (3) became attached to Playfair's name is a mystery. Henderson and Taimina suggested that T. L. Heath and Morris Kline were to blame, since Heath treated (3) and (2) as interchangeable in his 1908 translation of Euclid's *Elements* in the quotation below and Kline omitted (2) but described (3) as "Playfair's Axiom" and "the axiom used in modern books" (Kline 1972, p. 865):

> *Through a given point only one parallel can be drawn to a given straight line or, Two straight lines which intersect one another cannot both be parallel to one and the same straight line.*
> This is commonly known as "Playfair's Axiom," but it was of course not a new discovery. It is distinctly stated in Proclus' note to Eucl. I. 31 (Heath 1908, p. 220, emphasis in source).

While directing questioners to Henderson's and Taimina's work has sufficiently mollified them, ending the story there has never fully satisfied me. As a historian, I warn my students about assuming historical continuity between two points in time, and there is quite a distance between Playfair's 1795 publication date and Heath's 1908. Henderson and Taimina implied but did not follow up on the possibility that statements which addressed existence were introduced to rule out non-Euclidean geometries in research or perhaps even in teaching. Meanwhile, although historians of the parallel postulate and non-Euclidean geometry, at least as far back as Heath and Julian Coolidge (1940), have talked about "Playfair's Axiom," they have conflated (2) with (3), as we have already seen. (Besides Heath, fellow pioneer historian of mathematics Florian Cajori employed the term in (Cajori 1890, p. 378), but only with respect to (2).) Further, they have located Playfair and parallel axioms within that subject without considering their roles in the history of mathematics education, even though many references to "Playfair's Axiom" are in textbooks. Were 19th-century educators really connecting mathematicians' debates over whether the parallel postulate could be proven or the feasibility of non-Euclidean geometry to the use of the term "Playfair's Axiom" in teaching?

To try to clear up these loose ends, in 2007 I traced the application of the label to (2) back to Whewell's 1837 *Mechanical Euclid* (p. 151), but I was stymied by how "Playfair's Axiom" subsequently turned into (3). I have continued to watch for mentions of the term and periodically revisited digitized collections such as Google Books and the Internet Archive, which has unearthed both an earlier appearance of (2) being called "Playfair's Axiom" and evidence that (3) acquired the label by the late 1870s, about three decades earlier than previously known. On the surface, these searches for origins may appear to provide little more than a more accurate epilogue to a planned scholarly biography of Playfair. However, this article argues that the initial propagation of the term and its later pairing with a statement different in wording and logic transcend trivia to enhance our understanding of how audiences and priorities were in transition in geometry education in Great Britain and in the United States throughout the 19th century.

2 Parallels in *Elements of Geometry*

The eldest of eight children, in 1762 Playfair followed his father's footsteps to the University of St. Andrews to be trained as a minister in the Church of Scotland. He soon demonstrated interest and ability in mathematics and natural philosophy, but he failed to win the first two professorships for which he stood, in part because of his youth. He did succeed in making connections that allowed him to spend nearly four years in Edinburgh among Scottish Enlightenment figures, returning home after his father's death in 1772 to assume his father's parishes as well as financial responsibility for his mother and three sisters (Playfair and Jeffrey 1822; "Some Account" 1819). He maintained his intellectual contacts and tutored the sons of William Ferguson of Raith. When Adam Ferguson, apparently unrelated to William and the professor of mental and moral philosophy at the University of Edinburgh, suffered ill health in 1785, he exchanged positions with Dugald Stewart, the professor of mathematics. Playfair was hired to do the mathematics teaching. Since Ferguson drew the salary for their joint appointment, Playfair supported his family on the fees students paid to attend each lecture. He also continued to be active in the recently formed Royal Society of Edinburgh (Ackerberg-Hastings 2000).

Ten years into his employment, Playfair completed his seventh publication and first book, *Elements of Geometry; Containing the First Six Books of Euclid, with Two Books on the Geometry of Solids. To Which are Added, Elements of Plane and Spherical Trigonometry*. It is unclear when before his death in 1816, if ever, Ferguson stopped taking the joint chair's salary. Meanwhile, Playfair had also become responsible for his two nephews, one of whom would become the noted architect William Playfair. Besides his own financial needs that could be partially eased by book royalties, Playfair was confronted by issues with the existing geometry textbook, Simson's 1756 *The Elements of Euclid*, which consisted of Books I–VI and XI–XII that Simson had copyedited and clarified from Commandinus' 1572 Latin translation of Euclid's *Elements*. For instance, the woodcut plates for *The Elements of Euclid* had worn out over many printings, so the diagrams were gradually becoming illegible (Simson 1756; Burnett 1983). Additionally, Playfair was well aware of 18th-century developments in mathematics on the Continent. While he was generally a proponent of geometrical thinking, he also knew that analytical approaches had unleashed a tremendous amount of computing power that was underutilized by British mathematicians. He thus identified an opportunity to get his students to start recognizing connections between geometry and algebra (Ackerberg-Hastings 2002).

These financial, logistical, and intellectual aims came together in the preparation of *Elements of Geometry*. Playfair could expect that a new textbook would sell well in Edinburgh and elsewhere, since plane and solid geometry was a standard college course, valued for its preparation for further mathematics as well as for its mental training in proper Euclidean reasoning. A replacement for Simson with readable illustrations was in fact needed. And, although Playfair retained verbatim roughly 71% of Simson's overall content (363 definitions, postulates, axioms, and propositions with proofs and diagrams), he famously substituted symbols for words in Book V, which covered the theory of proportion. There, after reworking about one-third of the definitions but preserving the four axioms, Playfair kept almost all of Simson's propositions but rewrote—and, often, revised—every proof. For instance, in the demonstration for Proposition IV, Simson stated in prose:

> Let A the first have to B the second, the same ratio which the third C has to the fourth D; and of A and C let there be taken any equimultiples whatever E, F; and of B and D any equimultiples whatever G, H. E shall have the same ratio to G, which F has to H (Simson 1756, p. 144).

Playfair used symbols:

> Let A : B :: C : D, and let *m* and *n* be any two numbers; *m*A : *n*B :: *m*C : *n*D (Playfair 1795, p. 138).

Playfair also made large-scale changes to the content on solid geometry, Books XI and XII in *The Elements of Euclid* and Books VII and VIII in *Elements of*

Geometry, including adding approximately 18 new propositions and two definitions. Another five propositions were added to Book VI, on proportional relationships between plane figures. Analysis of all these revisions is outside the scope of this study but would be worth the attention of subsequent researchers.

While Playfair additionally demonstrated a willingness to tinker with the definitions throughout the text, he left alone all three of Simson's postulates and 14 of the 16 axioms (12 in Book I, four in Book IV). Simson had already dropped Euclid's fourth postulate, "All right angles are congruent," and moved the parallel postulate to the end of the axioms, where he enclosed it between quotation marks:

> If a straight line meets two straight lines, so as to make the two interior angles on the same side of it taken together less than two right angles, these straight lines being continually produced shall at length meet upon that side on which are the angles which are less than two right angles (Simson 1756, p. 7).

Playfair preserved the quotation marks but substituted statement (1):

> Two straight lines cannot be drawn through the same point, parallel to the same straight line, without coinciding with one another (Playfair 1795, p. 7).

Like Simson's, Playfair's typography seems to indicate acknowledgement of the long-known problematic nature of the parallel postulate. He went even further by introducing the replacement axiom, which changed the focus from the angles to the lines and considered the lines in relation to a point. This general approach to parallels had been around since at least the 5th century CE, when Proclus wrote *A Commentary on the First Book of Euclid's "Elements"*, although this particular wording appears not to have previously appeared in print. While a copy was not among Playfair's library when it was auctioned after his death, he might have known that Thomas Taylor had published a translation in London in 1792 (*Catalogue* 1820; Taylor 1792). Playfair surely knew that Proclus was a frequent topic in treatises on geometry or philosophy. Yet, he provided no indication of a source for the axiom; indeed, in the preface Playfair described (1) as "a new Axiom ... introduced in the room of the 12th, for the purpose of demonstrating more easily some of the properties of parallel lines" (Playfair 1795, p. vii). However, in the endnotes he made it clear that the second part of that sentence was more important to him than claiming novelty. There, he noted that (1) was "more obvious" than the parallel postulate (Playfair 1795, p. 354) and later explained that his chief goal for the opening of Book I was to provide statements and concepts that were readily grasped by students and avoided the need for sophisticated metaphysics:

> If I have not followed this method, it is because I wished to preserve the text of Euclid with the least alteration possible. I therefore assumed as an Axiom, a proposition that is not perhaps so obvious a property of straight lines as that which has just been stated, but which is certainly much more so than Euclid's, and one which I know from experience that beginners find no difficulty in comprehending, or in admitting to be true (Playfair 1795, p. 371).

In other words, he changed the parallel axiom so that it would be more easily understood by the teenaged boys who studied plane and solid geometry at the University of Edinburgh and elsewhere.

In 1804, Playfair made significant editorial changes for the second edition of *Elements of Geometry*, including converting the two books on solid geometry into a three-book supplement on the quadrature of the circle, the intersection of planes, and solid geometry. He also rephrased the parallel axiom and separated it from the other axioms—he gave the same numbered list of the first ten axioms and then, in the place of the eleventh axiom but still within quotation marks, provided this unnumbered sentence that we are calling statement (2):

> Two straight lines, which intersect one another, cannot be both parallel to the same straight line (Playfair 1804, p. 7).

Although numbering was restored to the eleventh axiom in the 1813 third edition, the last revision of the textbook to which Playfair contributed, Playfair preserved the wording of (2) there. (Reprints and editions after Playfair's death also contained (2).) He never revised his sentence about "a new axiom" in the preface (Playfair 1804, p. vi; Playfair 1813, p. vii), but in each edition he rewrote the extensive endnote for Proposition I.29. In 1813, he added that (2) "has been assumed by others, particularly by [William] Ludlam," in Ludlam's 1785 *The Rudiments of Mathematics* (Playfair 1813, p. 422). There, Ludlam recommended replacing the parallel postulate with:

> Two straight lines meeting in a point, are not *both* parallel to a third line (Ludlam 1790, p. 137, emphasis in source).

Note that (2) mirrors the logical structure of this statement but that its wording differs slightly. That Playfair finally acknowledged a source may not be significant. Since this attribution was buried in the endnotes—and was not even in a note for the eleventh axiom but rather was mentioned offhandedly within a lengthy note for a theorem in the last two-fifths of Book I—it likely was unread by most students and perhaps also overlooked by the authors who later popularized the term "Playfair's Axiom."

Although he changed the axiom in 1804 and added a citation in 1813, Playfair's motives for employing a substitute for the parallel postulate presumably remained the same, to ease the paths of teaching and learning. In the preface, he retained his point that alterations such as (1) or (2) would "in some degree facilitate the study of the Elements" (Playfair 1804, pp. v–vi; Playfair 1813, p. viii). He never revised his endnote comments on utilizing (1) after he replaced it with (2), which said that a parallel axiom "appeared more obvious" while the parallel postulate was "by no means self-evident" (Playfair 1804, p. 283; Playfair 1813, p. 410). Later in the endnotes, he was no longer as explicit about being motivated by a desire to help students more readily grasp the concept of parallel lines without having to address its underlying complexity. As his thinking on the theory of parallels evolved over time, he removed any mention of (1) or (2) from his endnote on proposition I.29 in the second edition and, in the third edition, in addition to citing Ludlam, noted only that he used (2) to permit a proof of I.29 that had "more brevity than Simson's" (Playfair 1804, pp. 288–289; Playfair 1813, p. 422). By mentioning that Playfair was discussing the theory of parallels (Playfair 1795, pp. 360–373; Playfair 1804, pp. 288–293; Playfair 1813, pp. 418–431), we can briefly acknowledge that he was indeed involved in ongoing debates about whether the parallel postulate could be proved—in part reworking the endnote on I.29 in order to incorporate what Adrien-Marie Legendre was saying in the endnotes to various editions of his own *Éléments de géométrie* (first edition published in Paris in 1794)—but suggest that readers consult (Ackerberg-Hastings 2000, 2002, 2007) for further information. Playfair's educational context was what influenced other educators to label (2) with his name.

3 From an Axiom in Playfair to "Playfair's Axiom"

By 1813, Playfair had used two different axioms for parallels and admitted that the underlying concept of both, along with the wording of (2), was not original to him. Yet, the statements were in editions of a textbook that carried Playfair's name, and so it would probably be unsurprising if users of the textbook associated the axioms with the source in which they read them. While Ludlam's *Rudiments of Mathematics* went through at least five printings and was used at St. John's and other Cambridge colleges for 30 years (Decesare 2011, pp. 112–113), its wider circulation paled in comparison to Playfair's *Elements of Geometry*. That text was printed at least 13 times in Great Britain between 1795 and 1875 as well as abridged or excerpted by numerous British editors, and it was adopted in other countries. Perhaps most notably, after Yale tutors taught elementary plane and solid geometry from an 1806 pirated edition of *Elements of Geometry*, it was reprinted 32 more times and used at colleges and academies throughout the United States (Vernau 1984, ciii:200–223, cclx:369; *National Union Catalog*, 1976, cdlxi:412–417; Karpinski 1940, pp. 163–165; Ackerberg-Hastings 2002, p. 69).

Such a well-known textbook thus almost inevitably was referred to as "Playfair's geometry" or simply "Playfair" in classrooms, college catalogues, student reminiscences, and textbook sellers' advertisements—such usages in the 19th century are indeed far too numerous to list. Similarly, the label "Playfair's Axiom" appears to have been affixed to (2) for educational purposes. Those who debated ideas about the nature of parallel lines in a research-like setting (which, admittedly, was often the endnote section of a geometry textbook), such as Legendre or John Leslie, did not mention Playfair by name even though he discussed their work directly (Leslie 1811, pp. 402–406; Legendre 1812, pp. 280–286; Playfair 1812; Playfair 1813, pp. 421–426). Reviewers and compilers of encyclopedias also paid attention to Playfair's treatment of Euclidean geometry, including his essays in the endnotes of *Elements of Geometry*. In 1798 an anonymous reviewer noted that Playfair gave (1) because he was dissatisfied with all previous efforts to resolve the problem that Euclid's parallel postulate is not self-evident (Review of *Elements of Geometry* 1798, pp. 159–160). The reviewer wrote out the statement but did not label it. In 1809, William Nicholson used Playfair's discussion of parallels as the basis for the beginning of his entry on "parallel," again noting that the fifth postulate in Euclid is not self-evident. He commented that Playfair "[introduced] a new axiom" but neither included it nor named it. Instead, he departed from *Elements of Geometry* by advocating for a replacement definition for "parallel," "Parallel straight lines are those whose least distances from each other are everywhere equal," and then proving five theorems about situations in which parallel lines exist (Nicholson 1809). Mathematicians and expository writers in the first quarter of the 19th century certainly did not coin the phrase "Playfair's Axiom." Educators were the ones who not only pulled out the statement (namely (2)) in order to highlight it in teaching but also *who* gave it a name.

One useful tool for locating the emergence and use of the appellation is the capability of online collections of digitized books for searching for terms and concepts, although search results can be inconsistent from one research session to another. At present, therefore, the primary sources available through sites such as GoogleBooks and HathiTrust suggest that (2) was called "Playfair's" at least as early as an annual examination administered in May 1830 at Queen's College, Cambridge:

> Apply Playfair's axiom, "Two straight lines which intersect one another, cannot be both parallel to the same straight line," to prove Prop. 29, Book I. "If a straight line fall upon two parallel straight lines, it makes the alternate angles equal to one another; and the exterior angle equal to the interior and opposite upon the same side; and likewise the two interior angles upon the same side together equal to two right angles" (Wright 1830, p. 31).

Whether this can definitively be called the first use of "Playfair's Axiom" is not as important as the reason for the naming. It appeared in a test because the students were expected to recognize it as a reference to their textbook. Later in the decade, William Whewell also applied the term to (2) in a textbook for mechanics, hydrostatics, and the laws of motion that he styled after Euclid's *Elements* and intended for use by students attempting examinations at Cambridge in the wake of reforms recommended by Whewell that re-emphasized geometrical methods in the "mixed mathematics" (Whewell 1837, pp. v–viii). After the main sections of the text, he made "remarks" about the nature of mathematical reasoning and induction that he hoped would help students remain generalists rather than "[pursue] too exclusively one particular line of mathematical study" (Whewell 1837, p. 146). In critiquing Dugald Stewart's preference for definitions over axioms, Whewell named "Playfair's axiom" as an example of a geometrical axiom that demonstrated the common logical ground for definitions and axioms (Whewell 1837, p. 151).

Some other representative examples from Great Britain can be provided. In 1853, Henry J. Hose used the phrase in *The Elements of Euclid . . . A New Text, Based on That of Simson* (p. 297). "Playfair's Axiom" was to be employed for teaching "the treatment of parallels" in the Association for the Improvement of Geometrical Teaching's 1871 recommendations for reforming British instruction; the report was reprinted in the *Quarterly Journal of Education* (1872, p. 166). In evaluating Edward Butler's 1872 *Supplement to the First Book of Euclid's Elements* and Francis Cuthbertson's 1874 *Euclidian* [*sic*] *Geometry* as alternatives to Euclid's *Elements of Geometry*, an anonymous reviewer said that the choice between the parallel postulate and "Playfair's axiom" was "merely a question of preference" (*London . . . Philosophical Magazine* 1874, p. 302). Generally, use of the term—almost always in conjunction with (2) rather than with (1)—gradually became more common in schools and colleges, with a dramatic upturn of occurrences in the last quarter of the 19th century that was accompanied by the twist in the story that is discussed in the next section.

4 The Mystery of the Substituted Axiom

The story of "Playfair's Axiom" did not end with the pairing of the label and statement (2). As was noted in the introduction, today many people associate the term with some variation of the axiom we have called statement (3):

> Given a line and a point not on the line there is one and only one line through the point that is parallel to the given line.

The phrasing differs from (1) and (2) both in its semantics, by starting with a line and a point rather than with two lines, and in its logic. Unlike the statements in Playfair's textbooks, which claimed only the uniqueness of a parallel line (if there is a line parallel to the given line, there is no more than one), (3) posits both uniqueness and existence (certainly there is at least one line parallel to the given line). So, the "Playfair's Axiom" that today's students and professors associate with him was not only never invented by John Playfair, it was never stated by him as an axiom in his geometry textbook. And, it asserts more than either of the axioms that he did think had educational value, permitting (3) to rule out non-Euclidean spaces in the study of Euclidean geometry.

Versions of (3) had circulated for centuries, but they were probably employed most frequently as theorems to be proven rather than as self-evident axioms or postulates. For instance, in the same encyclopedia entry in which he compared and contrasted the parallel postulate and (2), Nicholson suggested that (3)—albeit in a form that merely asserted uniqueness—could be viewed as a corollary to the theorem "If two straight lines be perpendicular to the same straight line, they are parallel to each other": "Hence it appears, that through the same point no more than one line can be drawn parallel to the same straight line" (Nicholson 1809). In 1812, Legendre prepared a proof of the same concept, "Par un point donné A on ne peut mener qu'une seule parallèle à une ligne donnée BD," as a corollary to his theorem I.22 in the ninth edition of *Éléments de géométrie* (Legendre 1812, p. 23). That same year, in a review of the second edition of Leslie's *Elements of Geometry*, Playfair entertained the possibility that (3) could serve as an axiom rather than as a provable statement, although he never mentioned the matter again:

> The work before us adds one, in our opinion, to the number of the failures [to prove a converse of the Parallel Postulate, that parallel lines form equal alternating angles], though the author seems completely satisfied of the contrary. It must be remarked, that if it were proved that *through the same point there cannot be drawn more lines than one parallel to a given line* (a proposition that may perhaps be assumed as an *axiom*), we should be in possession of a criterion equivalent to that which has just been mentioned; and the doctrine of parallel lines would no longer have any difficulty (Playfair 1812, p. 88, emphasis added).

Leslie had eliminated all postulates and axioms and indeed seems to have tried to avoid the problems associated with parallels as much as possible—his definition stated merely, "Straight lines which have no inclination, are termed *parallel*" (Leslie 1811, pp. 7, 406, emphasis in source). He transformed Euclid's Proposition I.29 into a corollary that contained another way of stating a uniqueness-only version of (3):

> Since the position CD is individual, or that only *one* straight line can be drawn through the point F parallel to AB, it follows that the converse of the proposition is likewise true, and that those three properties of parallel lines are *criteria* for distinguishing parallels (Leslie 1811, p. 27, emphasis in source).

By 1855, (3) had moved from supplemental notes and attempted proofs to full status as a postulate or axiom in English-language plane and solid geometry textbooks, although no evidence has yet been found to explain why educators turned away from (1) or (2) as well as Euclid's parallel postulate. James B. Dodd, a mathematics and natural philosophy professor in Kentucky, gave an existence-only phrasing in *Elements of Geometry and Mensuration*:

> A POSTULATE requires something to be granted as *a fact*, or as the result of an operation, the *possibility* of which is self-evident; such as. . . .
> That through a given point a straight line has been drawn parallel to a given straight line (Dodd 1855, p. 6, emphasis in source).

Other than mentioning that he collected materials "from several sources besides the well known treatises of *Euclid* and *Legendre*," he offered no explanation for making the change and actually went on to demonstrate the construction of such a parallel line (Dodd 1855, pp. iv, 100, emphasis in source). The text was apparently only printed in one run, so it likely was not influential in the United States and surely was unknown in Great Britain. Another phrasing of statement (3) that only asserted uniqueness was published as one of the "principles of construction" in Holloway's 1865 *Mental Geometry*: "Through a given point without a given straight line, one parallel line may be drawn" (Holloway 1865, pp. 3, 56). By the 1890s, (3) was commonly appearing as a postulate or axiom in American school geometry textbooks without any naming or label.

While Americans were thus starting to experiment with alternative postulates or axioms, just as they were more broadly undertaking a variety of new approaches in teaching and textbooks that they gave names such as "inductive" or "mental" or "heuristic," British educators seem to have lagged at least a decade behind in considering (3) as a replacement for the parallel postulate and (2). In 1869 in South Kensington, William Kingdon Clifford delivered ten lectures on geometry titled "On Plane Surfaces and Straight Lines" and aimed at women. The syllabus for the lectures consisted of a list of statements—definitions, axioms, and theorems—presumably explained and discussed by Clifford, followed by five homework-like questions. The eleventh item read:

> Only one straight line can be drawn parallel to a given straight line through a given point. It is then parallel to the given line at all points on it (Clifford 1882, p. 629).

By 1878, for reasons that remain unknown, English educators had decided to pair (3) with Playfair's name despite surely knowing that his *Elements of Geometry* did not state it—recall that book was still in print:

> To obviate this difficulty [that the truth of the parallel postulate is not obvious], different axioms have been proposed by some writers on geometry, while others have attempted to establish a theory of parallels without any axioms; on all these attempts I need only remark that they have all been found unsatisfactory. The only axiom I shall notice is that given by

> Playfair and many subsequent writers: "Through a given point only one straight line can be drawn parallel to a given straight line" (Buchheim 1878, pp. 78–79).

At the end of the paper, the author, Arthur Buchheim, referred to the statement with the label "Playfair's axiom" in arguing that it was equivalent to the parallel postulate and so also not self-evident. Rather, he saw these statements and Euclid's ninth axiom ("The whole is greater than its part") as "principles that determine the geometry of the plane" (Buchheim 1878, p. 81). Buchheim was a recent graduate of the City of London School, an independent day school in Cheapside that was known in the 19th century for admitting boys regardless of their religion and for offering an excellent science and math education, as evidenced by the performance at Cambridge or Oxford by students turned faculty such as Edwin Abbott Abbott, the author of *Flatland*. Indeed, Buchheim had earned a mathematical scholarship of £100 for five years to New College, Oxford (A.A 1877). He was also well-versed in all of Clifford's mathematics, and in fact he helped propagate Clifford's ideas about reforming geometry teaching in England, which provided the larger context for his interest in the structures of various geometries (Tattersall 2005, pp. 201–202). It is possible that the decision to replace either of the axioms in Playfair's *Elements of Geometry* with (3), which could account for the existence of non-Euclidean geometry by adding verbiage to stipulate the existence of parallel lines, arose out of the Clifford circle's project to use projective geometry as a means for thinking about non-Euclidean geometry. (See, for example, (Richards 1988).) In the United States 27 years later, P. A. Lambert thought about (3) in exactly this way, although he did not associate the statement with Playfair's name. He noted that, if the axiom of parallels was phrased "Through a given point without a given line one and only one parallel to the line can be drawn," then Euclidean geometry resulted. Changing the axiom to allow two parallels to exist generated hyperbolic geometry, and replacing "one and only one" with "no parallel" produced elliptic geometry (Lambert 1905, p. 83).

Once Buchheim put the statement and label together, it did not take long for other mathematicians and educators to follow suit. In particular, the pairing proliferated after the turn of the 20th century. For instance, (3) was called "Playfair's Axiom" at least three times in 1903. In *Elementary Geometry: Practical and Theoretical*, Charles Godfrey and Arthur Warry Siddons simply inserted the statement about one-fifth of the way through Book I of "theoretical geometry" (Godfrey and Siddons 1903, p. 74). Similarly, Slovenian mathematician Franz Hocevar—or his English editors—mentioned it in a parenthetical comment suggesting that Playfair's Axiom held both on a plane and in three-dimensional space (Hocevar 1903, p. 2). Meanwhile, Rupert Deakin, headmaster of King Edward's Grammar School in Stourbridge, claimed that (3) and (2) could both be called "Playfair's Axiom" in a self-study textbook offered by University Correspondence College in London, stating in part:

> Many attempts have been made to improve on this part of Euclid [the 12th axiom], but none are satisfactory. The simplest axiom that has been suggested is the following, which is known as Playfair's Axiom:—

> *Through a given point only one straight line can be drawn parallel to a given straight line.* This is the same as:— *Two straight lines which intersect cannot be both parallel to the same straight line.* In this form Playfair's Axiom is the converse of Euclid's Prop. 30, and has therefore been objected to (Deakin 1903, p. 70, emphasis in source).

Perhaps this conflation, along with Heath's five years later, helped ensure that (2) would be replaced by (3). Despite a brief period of overlap in which some authors applied "Playfair's Axiom" to (2) and some used the label to refer to (3), the overwhelming majority of authors who wrote about "Playfair's Axiom" in the 20th and 21st centuries stated it as (3). Like these three books, almost all of the sources that brought up "Playfair's Axiom" were textbooks, with some works aimed at popular audiences. Educators also used the term when they wrote notes on teaching and book reviews for each other, indicating that it had taken root in school instruction. For example, after two occurrences in the English periodical *The Mathematical Gazette* in 1898, "Playfair's Axiom" was discussed in 13 articles in the first decade of the 20th century, six pieces in the 1910s, 17 in the 1920s, and eight in the 1930s. Thus, even though (3) on its own may have gained some popularity in Great Britain when it was articulated by a professional mathematician such as Clifford, the pairing of label with concept and subsequent widespread usage occurred in educational contexts, just as with the first dissemination of the term in the middle third of the 19th century.

5 Conclusion

For many mathematicians and mathematics educators, "Playfair's Axiom" is the only thing they know about John Playfair. Unfortunately, there are problems with the attribution that go beyond a simple application of Stigler's Law of Eponomy, that Playfair's name was attached to the axiom even though he did not invent it. Playfair published two different phrasings of an axiom of parallels, (1) in 1795 and (2) in 1804; in 1813 he attributed (2) to William Ludlam's *The Rudiments of Mathematics*. In both cases, he substituted out Euclid's parallel postulate because he thought the alternatives were easier for students to understand—he was not attempting to find a statement that would enable mathematicians to prove the parallel postulate or be accepted by them as truly self-evident. Other educators presumably found Playfair's wording helpful in their classrooms, as British students found references to "Playfair's Axiom" on their examination papers no later than 1830.

Meanwhile, statements resembling (3) moved from theoretical discussions of problems with the parallel postulate into geometry teaching, particularly in the United States. By 1878 in England, these existence-only or uniqueness-only versions of (3) were rewritten to account for both uniqueness and existence and paired with Playfair's name, again for educational purposes and coinciding with attempts to reconcile non-Euclidean geometry with the standard subject of Euclidean geometry so that both could be taught in English schools. More broadly, educators on both sides of the Atlantic Ocean increasingly advocated in textbooks and educational journals for "Playfair's Axiom" to replace the parallel postulate. In the decades around 1900, it is not always clear whether a particular author was thinking of (2) or (3). Yet, although a definitive answer for why it happened may not be possible, early in the 20th century (3) almost entirely supplanted (2) as what was meant by references to "Playfair's Axiom." As more 19th-century books and periodicals are digitized, perhaps later researchers will be able to move these dates even closer to the 1795 initial publication of *Elements of Geometry*. Nonetheless, nailing down these "famous firsts" is not as valuable to historians as trying to make sense of the larger context of how "Playfair's Axiom" fits within the story of secondary and undergraduate mathematics education.

Historians have generally only thought about Playfair and parallels as part of the chronology of attempts to prove Euclid's parallel postulate: significantly after Saccheri and Lambert, roughly concurrent with the several proofs found in the several editions of Legendre's geometry textbook, and decades before Lobachevsky and Bolyai proposed non-Euclidean geometries that did not require a parallel postulate. However, while Playfair engaged with questions about meaning or other philosophical hand-wringing that characterize the standard chronology of the parallel postulate as one facet of his overall career, he separated that theorizing from his decisions about which axioms to publish. What is striking about the emergence of "Playfair's Axiom" as a term is that it also was coined strictly for addressing needs that arose in teaching undergraduate and school geometry. As we have followed the meandering path required to fully answer the questions I hear at mathematics conferences, I hope readers have learned something new about how plane and solid geometry were taught in English-speaking areas in the 19th century as well as about when the label was coined and how it came to be transferred to a logically different statement. Perhaps this twisted tale can additionally serve as a reminder to ask the right questions about how and why certain concepts or terms became commonplace, even when they were misnamed or fundamentally transformed. The history of mathematics education, for one, contains numerous such episodes waiting to be explained.

Acknowledgements The author thanks the anonymous referees and audiences at the 2016 Joint Mathematics Meetings and CSHPM Annual Meeting for their comments. April 2017 attendees of the Philadelphia Area Seminar on the History of Mathematics helped workshop revisions to the initial paper.

References

A. A. (1877) School News. The City of London School Magazine 1, no. 7:189.

Ackerberg-Hastings Amy (2000) Mathematics is a Gentleman's Art: Analysis and Synthesis in American College Geometry Teaching, 1790–1840. Ph.D. diss., Iowa State University.

Ackerberg-Hastings Amy (2002) Analysis and Synthesis in John Playfair's *Elements of Geometry*. British Journal for the History of Science 35:43–72.

Ackerberg-Hastings Amy (2007) Euler and the Enlightenment Mathematicians: A Scottish Perspective. Proceedings of the Canadian Society for History and Philosophy of Mathematics, vol. 20, 12–23. Thirty-Third Annual Meeting, 27–29 July.

Association for the Improvement of Geometrical Teaching (1872) The Improvement of Geometrical Teaching. Quarterly Journal of Education and Scholastic Advertiser n.s., no. 3:164–168.

Buchheim Arthur (1878) Euclid's Tenth and Twelfth Axioms. The City of London School Magazine 2, no. 14:77–81.

Burnett John (1983) Robert Simson's Euclid and the Foulis Press. Bibliotheck 11:136–148.

Cajori Florian (1890) The Teaching and History of Mathematics in the United States. Washington, DC: Government Printing Office.

Catalogue of the Library of the Late John Playfair, Esq (1820) Edinburgh: James Ballantyne and Co.

Clifford William Kingdon (1882) Syllabus of Ten Lectures to Ladies on Geometry Delivered at S. Kensington [1869]. In: Tucker Robert (ed) Smith H. J. Stephen (intro) Mathematical Papers by William Kingdon Clifford. London: Macmillan and Co.

Coolidge Julian Lowell (1940) A History of Geometrical Methods. Oxford: Clarendon Press.

Deakin Rupert (1903) Euclid: Books I., II. 2nd edn. The University Tutorial Series. London: University Tutorial Press.

Decesare Richard (2011) William Ludlam: Portrait of an Eighteen-century Mathematician. BSHM Bulletin 26:105–117.

Dodd James B (1855) Elements of Geometry and Mensuration. New York: Farmer, Brace & Co.

Godfrey Charles, Siddons Arthur Warry (1903) Elementary Geometry: Practical and Theoretical. Cambridge: University Press.

Heath Thomas L (trans, intro) (1908) The Thirteen Books of the Elements. Euclid. Vol. 1. Cambridge: Cambridge University Press.

Henderson David W, Taimina Daina (2005) How to Use History to Clarify Common Confusions in Geometry. In: Shell-Gellasch Amy, Jardine Dick (eds) From Calculus to Computers: Using the Last 200 Years of Mathematics History in the Classroom. MAA Notes No. 68. Washington, DC: Mathematical Association of America.

Hocevar Franz (1903) Solid Geometry. Godfrey C, Price EA (trans, eds) London: Adam and Charles Black.

Holloway Henry H (1865) Mental Geometry, or Generalizations of Geometrical Demonstrations in Planes, Solids, and Spherics. Philadelphia: J. B. Lippincott & Co.

Hose Henry J (1853) The Elements of Euclid . . . A New Text, Based on That of Simson. London: Whittaker and Co.

Karpinski Louis C (1940) Bibliography of Mathematical Works Printed in America Through 1850. Ann Arbor.

Kline Morris (1972) Mathematical Thought from Ancient to Modern Times. Oxford: Oxford University Press.

Lambert P A (1905) The Straight Line Concept. Proceedings of the American Philosophical Society 44:82–89.

Legendre A M (1812) Éléments de géométrie, avec des notes, 9th edn. Paris: Firmin Didot.

Leslie John (1811) Elements of Geometry, Geometrical Analysis, and Plane Trigonometry, 2nd edn. Edinburgh: James Ballantyne and Co.

Ludlam William (1790) The Rudiments of Mathematics; Designed for the Use of Students at the Universities: Containing an Introduction to Algebra, Remarks on the First Six Books of Euclid, the Elements of Plane Trigonometry, 3rd edn. London: J. Nichols.

Mazur Joseph (2006) Euclid in the Rainforest: Discovering the Universal Truth in Logic and Mathematics. New York: Plume.

The National Union Catalog: Pre-1956 Imprints (1976) 753 vol. London.

Nicholson William (1809) The British Encyclopedia, or Dictionary of Arts and Sciences. Vol. 5. London.

[Playfair James G, Jeffrey Francis] (1822) Biographical Memoir. In: Playfair James G (ed) Works of John Playfair, i:ix–cv. Edinburgh.

Playfair John (1795) Elements of Geometry; Containing the First Six Books of Euclid, with two books on the geometry of solids. To Which are Added, Elements of Plane and Spherical Trigonometry. Edinburgh: Bell & Bradfute.

Playfair John (1804) Elements of Geometry, Containing the First Six Books of Euclid, With a Supplement on the Quadrature of the Circle and the Geometry of Solids, 2nd edn. Edinburgh: Bell & Bradfute.

[Playfair John] (1812) Review of Elements of Geometry, Geometrical Analysis, and Plane Trigonometry, 2nd edn, by John Leslie. Edinburgh Review 20:79–100.

Playfair John (1813) Elements of Geometry; Containing the First Six Books of Euclid, With a Supplement on the Quadrature of the Circle and the Geometry of Solids, 3rd edn. Edinburgh: Bell & Bradfute.

Review of Elements of Geometry, by John Playfair (1798) Monthly Review n.s., 26:154–165.

Richards Joan L (1988) Mathematical Visions: The Pursuit of Geometry in Victorian England. San Diego: Academic Press, Inc.

Simson Robert (1756) The Elements of Euclid, viz. the First Six Books, together with the Eleventh and Twelfth. Glasgow: Robert and Andrew Foulis.

Some Account of the Character and Merits of the Late Professor Playfair (1819) Edinburgh Magazine and Literary Miscellany August:163–168.

Stein, James D (2008) How Math Explains the World: A Guide to the Power of Numbers, from Car Repair to Modern Physics. New York: HarperCollins.

Suzuki Jeff (2002) A History of Mathematics. Upper Saddle River, NJ: Prentice Hall.

Vernau Judi (ed) (1984) The British Library General Catalogue of Printed Books to 1975. 360 vol. London and Munich.

Tattersall Jim (2005) Arthur Buchheim: Mathematician of Great Promise. Proceedings of the Canadian Society for History and Philosophy of Mathematics, vol. 18, 200–208. Thirty-First Annual Meeting, 4–6 June.

Taylor Thomas (trans. and ed.) (1792) The Philosophical and Mathematical Commentaries of Proclus on the First Book of Euclid's "Elements". London.

Whewell William (1837) The Mechanical Euclid, Containing the Elements of Mechanics and Hydrostatics Demonstrated After the Manner of The Elements of Geometry; and Including the Propositions Fixed Upon By the University of Cambridge as Requisite for the Degree of B.A. Cambridge: J. and J. J. Deighton.

Wright J M F (1830) A Collection of Cambridge Mathematical Examination Papers; As Given at the Several Colleges. Part I: Containing Papers in the Branches of the Pure Mathematics. Cambridge: W. P. Grant.

Napier, Torporley, Menelaus, and Ptolemy: Delambre and De Morgan's Observations on Seventeenth-Century Restructuring of Spherical Trigonometry

Joel S. Silverberg

Abstract An effort to reorganize and systematize planar and spherical trigonometry began in the 15th century with the work of Regiomontanus, extended throughout the 16th century with work by Otho, Rheticus, Pitiscus, and Fincke, and continued into the 17th century by Napier, Torporley, Viète, and others. During the 18th and 19th century, publications by Euler, Taylor, Fourier, and Gauss extended the role of trigonometric functions into new areas including power series, and complex functions of complex variables. An analysis of De Morgan's criticism of Napier's and Torporley's efforts in this area sheds light on the challenges to an historian of mathematics of one era attempting to understand the thought process of mathematicians living in earlier times. In particular, we focus on two areas: the historian's knowledge of future mathematical developments and modes of expression unknown to those living in the earlier period, and secondly an incomplete, inaccurate, or absent knowledge on the part of the historian of definitions, references, or conventions well known to those of the earlier era. These definitions, references, and conventions were often used without comment or explanation, and occasionally used without mention, since the writer could assume them to be common knowledge to the readers of his time, and that their use would be understood by his readers, even if that use was implicit.

Keywords Early modern • Seventeenth-century • Spherical trigonometry • Menelaus figure • Quadrantal triangles • Torporley • Napier • Delambre • DeMorgan • Triplicity

J.S. Silverberg (✉)
Professor Emeritus, Roger Williams University, Providence, RI, USA
e-mail: joel.silverberg@alumni.brown.edu

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*,
Proceedings of the Canadian Society for History and Philosophy of Mathematics
La Société Canadienne d'Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_10

1 Introduction

In this paper I revisit a controversy I first encountered while researching two early seventeenth-century attempts to simplify the rules for solving right-angled spherical triangles: Napier's *Pentagramma Mirificum* (Silverberg 2008) and Toporley's Bishop's Mitres with their twin portals : Dextra and Sinistra and their twin sets of tables : Quadrans and Quincunx . (Silverberg 2009)

Napier's contemporaries, including Briggs, Wright, Brahe, and Kepler, thought very highly of his mathematical contributions, yet the general opinion during the mid nineteenth through mid twentieth century was that his circular parts were little more than a clever mnemonic device or computational aid, but of no mathematical interest. The opinion was also generally voiced that Napier left no clues as to where these rules came from. A handful of mathematicians voiced the contrary opinion that these rules were important theoretical contributions to spherical trigonometry and deserved to be treated and understood as theorems. The two "camps" of opinion on Napier's rules generally agreed that Napier's explanation of how they were developed is at best incomplete and often characterized as somewhere between obscure and altogether absent.

Augustus De Morgan published at least three different articles on the questionable value of each of these approaches and further argued that it was very likely that Napier had seen Torporley's work and that the credit for Napier's "rules of circular parts" more properly belonged to Torporley. The first of these articles appeared in 1838.

After describing the use of Napier's Rules, which condense the problem to the application of one of two simple rules: the sine of the middle part is either the product of the tangents of the adjacent parts or the product of the cosines of the opposite parts De Morgan states[1]:

> But we should strongly recommend the student to have nothing to do with this artificial memory, for it involves a process upon every occasion; and while one person is learning which are the parts, which have complements taken, and the rules, another will master the six results, and will have no occasion for any future process. These results are :
>
> 1. Cosine of hyp. = product of cosines of sides.
> 2. Cosine of hyp.= product of cotangents of angles.
> 3. Sine of side = product of the sine of hyp. and the sine of the opposite angle.
> 4. Tang. of side = product of the tang. of hyp. and the cos. of the adjacent angle.
> 5. Tang. of side = product of tang. opposite angle and sine of the other side.
> 6. Cos. of angle = cos. of opposite side and sine of the other angle.
>
> These pairs present analogies which will help the memory, and we should recommend them in preference to the rules of circular parts.(De Morgan 1837)

[1] Article on Circular Parts (Napier's), page 195.

2 De Morgan on Napier's Rules of Circular Parts : A Controversy Is Born

The Penny Cyclopædia article on Circular Parts was followed six years later by two additional communications from De Morgan: one an article for the Penny Cyclopædia on Vieta (De Morgan 1843b); the second a three page letter to the editors of the Philosophical Magazine and Journal (De Morgan 1843a). This letter was published under the title: *On the Invention of the Circular Parts*, and in it he claims that Napier's work on circular parts was taken from Nathaniel Torporley's 1602 publication *Diclides cœlometric*æ a full twelve years before Napier's *Mirifici logarithmorum canonis descriptio.* (Napier 1614b)

To the Editors of the Philosophical Magazine and Journal.

GENTLEMEN,

The object of this communication is to point out that a very material portion of Napier's "rule of circular parts" was published by Torporley twelve years before it was published by Napier. I do not mean that Torporley gave his theorems in as elegant a form as Napier, nor do I deny that Napier abbreviated Torporley: what I say is that, of that abbreviation which is usually attributed entire to Napier, a considerable part belongs to Torporley Nor am I by any means convinced that Napier had seen Torporley's work: but, according to the rule established in such cases, the *first* publisher must take precedence of all subsequent ones, whether independent discoverers or not.

...

The question whether Napier had seen Torporley's work cannot easily be settled: he uses the word *triplicity,* which is one of frequent occurrence in Torporley, and the figure of his demonstration contains the three triangles put together in exactly the same way as Torporley's *mother and daughters* come together on the *mitre."* These circumstances are not conclusive, but they are suspicious: *triplicitas* was by no means a common word among mathematicians; it was a technical term of judicial astrology, and would probably be avoided by the geometer: Torporley was an astrologer, as appears from the opening of his work. *Trias* and *ternio* would suggest themselves first (*trinitas* being excluded for an obvious reason). On this word probably the question will turn: if it should be found that no mathematicians of the period use the word *triplicitas* except Torporley and Napier, it will be difficult to avoid presuming that the latter must have seen the work of the former.

I remain, Gentlemen,

Yours faithfully,

University College, March 13, 1843 A De Morgan.

3 De Morgan on Torporley

At the end of a lengthy article on Viète for the Penny Cyclopædia [listed under the Anglisized form: Vieta], De Morgan begins a section on "minor mathematicians" in Viète's circle.

This article is the proper place of reference to several minor mathematicians, who are hardly worth separate articles in any except a very full biographical dictionary, but who owe some

> of their fame to their connection with Vieta. We may instance Nathaniel Torporley, Adrian van Roomen, Marino Ghetaldi, and Alexander Anderson.
> (De Morgan 1843b)

After providing a brief biography of Torporley, including his studies at Oxford, his years in France as amanuensis to Vieta, and his connections with Harriot and the Earl of Northumberland, De Morgan turns to Torporley's work.

> Torporley afterwards wrote his 'Diclides Coelometricae, sue Valvae Universales,' &c., London, 1602, and other works which we have never seen. In looking through the 'Diclides,' &c., which is mostly on spherical trigonometry . . . we found to our surprise, that Torporley had preceded Napier by twelve years in the publication of the greater part of the rule of Circular Parts, not indeed in Napier's convenient form, but with a complete reduction of the six cases to two, and rules, such as they were, by which to assimilate the connected cases. For more account of Torporleys process, which is the greatest burlesque on mnemonics we ever saw, we refer to the 'Philosophical Magazine' for May.
> (De Morgan 1843b)

De Morgan's letter to the Philosophical Magazine reveals more details of how he came to encounter the work of Torporley.

> An account of Torporley may be seen in Anthony Wood's *Athenae Oxonienses,* or in the article *Vieta* in the Penny Cyclopaedia.[2] An account of his work is in Delambre's *Astronomie Moderne,* vol. ii. p. 36. It is very strange that Delambre should not have seen that the very description which he gives almost amounts to stating Napier's rules: but it is to be remembered that these circular parts, so celebrated in Britain, have hardly ever been used abroad. Delambre himself (Astronomy, vol i. p. 205) says that he prefers to remember the six equations at once: a preference in which I heartily concur.
> There never was, perhaps, a more ludicrous mnemonical attempt than that of Torporley. . . . Torporley has given two tables of double entry, which Delambre says are the most obscure and incommodious that ever were made. The first is neither one nor the other; a and b being the arguments, and c the tabulated result, it amounts to $\tan c = \tan a \times \sin b$, the double entry being contrived like that of the common multiplication table. Of the second table, as the book is scarce, I subjoin half-a-dozen instances.

	30°			
	G	G	M	M
24°	21		27	
		8		33
	12		54	

	54°			
	G	G	M	M
70°	27		55	
		26		5
	1		50	

	70°			
	G	G	M	M
54°	39		14	
		30		46
	8		28	

	90°			
	G	G	M	M
19°	71		58	
		18		2
	53		56	

	19°			
	G	G	M	M
90°	9		30	
		9		30
	0		0	

	147°			
	G	G	M	M
59°	88		3	
		58		57
	29		6	

[2]Note the self references in these two articles.

> As far as the formulæ for right-angled triangles are concerned, this table applies as follows. The sine of the angle on the left, multiplied by the sine of the upper angle in the square compartment, gives the sine of the second angle in that compartment. Thus Torporley means to say that
>
> $$\sin\ 24^\circ \times \sin\ 21^\circ 27' = \sin\ 8^\circ 33'.$$
>
> Those who like such questions may find out the meaning of the other parts of the table.(De Morgan 1843a)

De Morgan shows more than a hint of frustration here and has clearly failed to decipher much of what is going on in this table. If the tables were computed by multiplying sin a by sin b to determine sin c, Torporley would have repeated the structure of the first table (named Quadrans), placing the arguments at the top and left margins of the table and the tabulated result placed in the body of the table. The table named Quincunx would have a column for every degree of arc, as does the Quadrans. Instead, Quincunx requires entry for a to select a row of the table, but the entry for b lies not in the column heading (as with Quadrans) but to be the uppermost of the three values in a cell within the body of the table (expressed in what seem to be unpredictable arcs expressed in degrees and minutes). The value of c is found in the middle of the three values in the same cell of the table which held b. But what are the meanings of the column headers and of the third entry of the set of three values in each cell of the body of the table? Such questions go to the heart of the purpose and means of construction of these tables, and deserve more than the dismissive comment that "those who like such questions may find out the meaning of the other parts of the table." Even the very names that Torporley has chosen for his two tables are somewhat mysterious. Some of the issues that this presents I will discuss in the following sections; others I intend to address in a future presentation or article.

4 De Morgan and Delambre

It is apparent, then, that De Morgan has discovered Torporley and his work through Book VII of the second volume of Delambre's *Astronomie Moderne* which is devoted to the work of Rheticus, Pitiscus, and Briggs in developing trigonometric tables for astronomical use. Toward the end of this chapter, Delambre discusses the work of Adrien Romain, Nathaniel Torporley, Philippe Lansberge, Christopher Clavius, Henry Briggs, Henry Gellibrand, Nathaniel Roe, William Oughtred, and Henry Sherwin. (Delambre 1821) A close reading of Book VII suggests that De Morgan was strongly influenced by Delambre's exposition. Although Delambre does not notice a close affinity between Napier's work and Torporley's he does view Torporley's work as bizarre in the extreme, and Napier's circular parts as a parlour trick without much value, an opinion echoed by De Morgan.

4.1 Delambre on Rheticus, Otho, Romanus, and Torporley

> Adrien Romain, dont nous avons párle á l'article de Viète ... publia en 1609 un ouvrage sous ce titre: *Triangulorum sphaericorum brevissimus simul ac facillimus quam plurimisque exemplis optie projectis, illustratus in gratiam Astronomiae, Cosmographiae, Geographiae, et Horologiographiae Studiosorum, jam primum editus, etc.*
> Effrayé de l'horrible prolixité de Rhéticus et d'Othon, il réduisit toute la Trigonométrie sphérique à six problèmes, dont tous les autre ne sont que des cas particuliers. Il goûte peu le moyen des perpendicularires qui partagent un triangle quelconque en deux rectangles; il préférait les pratique indiquées par Viète; mais ...
> Malgré ces promesses manifiques, les six problèmes, présentés en un nombre de classes qui monte jusqu'a dix-sept, forment un tout aussi effrayant et presque aussi difficile à compredre que le fatras de Rhéticus.[3]
> A la suite de cet ouvrage obscur et singulier, l'ordre des temps en amene un de meme genre, qui n'a cependant avec le premier aucun trait de ressemblance que la bizarrerie; c'est celui que Torporlaeus a publié sous ce titre: *Diclides Coelometricae ... Londini, 1602.*[4]

4.2 Consequences of De Morgan's Letter

De Morgan's repeated statements that Napier's famous rules were nothing more than a device to avoid the memorization of identities, and that further they were a hindrance to the development of a good mathematical memory, both ideas he gleaned from Delambre, left a lengthy trail.

A review of the literature reveals a flurry of interest in this claim, lasting nearly a century (between 1840 and 1940). Napier's contemporaries, including Briggs, Wright, Brahe, and Kepler thought very highly of his mathematical contributions, yet the general opinion during the mid-nineteenth through mid-twentieth century was that his circular parts were little more than a clever mnemonic device or computational aid, and of no mathematical interest.

Although the number of mathematicians and historians of mathematics who participated in this debate was modest, it included more than a few persons of note. I provide here a chronological sampling of comments upon Napier's Rules.

---

[3] Adrien Romain, of whom we have spoken in the article on Viète ... published in 1609 a work under the title *Triangulorum sphaericorum brevissimus*. Frightened by the terrible prolixity of Rheticus and Otho, he reduced all of Spherical Trigonometry into six problems, of which all other problems were special cases. He avoids having anything to do the perpendiculars which divide any triangle in to two right angled triangles. He prefers the methods pioneered by Viète, but ... In spite of these magnificent promises, the six problems, presented in a number of classes that mount up to no fewer than seventeen, form a whole which is every bit as frightening and nearly as difficult to understand as the rubbish of Rheticus.

[4] And following this most obscure and singular work, we come at last to another work of the same type, with which it shares nothing except its eccentricity. It is that which Torporley published under the title *Diclides Coelometricae ... London, 1602.*

- “There is no separate and independent proof of these rules; but the rules will be manifestly just, if it can be shewn that they comprehend every one of the ten results.” [Robert Woodhouse (holder of the Lucasian chair 1820–1822)]
- “These rules are proved to be true only by shewing that they comprehend all the equations which we have just formed.” [Sir George Biddell Airy, Lucasian Professor of Astronomy 1826–1828, Astronomer Royal 1835–1881]
- Napier’s rules are “mnemonical formulas” and “only create confusion instead of assisting the memory.” [Augustus De Morgan, 1843]
- “By putting these ten rules under a different form, Napier contrived to express them all in two rules, which, though artificial, are very generally employed as aids to the memory.” [William Chauvenet (1850)]
- “Napier’s rules for the solution of right-angled spherical triangles are generally presented merely as a memoria technica; and when so presented do not exhibit the principle upon which they depend.” [R.L. Ellis (1863)]
- “They are often expressed in textbooks on spherical trigonometry as if they were mere mnemonics, and have been thus regarded by men like Airy and De Morgan, who, one would have expected, might have appreciated their proper setting.” [D M Y Sommerville (1914)]
- “It is highly fitting that [Napier’s] rule for the circular parts should be rescued from the rubbish heap of mnemotechnics and be assigned its proper place as the most beautiful theorem in the whole field of elementary trigonometry.” [Robert Moritz (1915)]

5 Historiographical Challenges

How do we make sense of the contributions of a man who was admired and praised by the like of Briggs and Kepler, but whose contributions were considered trivial by mathematicians and astronomers such as Delambre, De Morgan, Chauvenet, and Airy? How do we make sense of Delambre’s dismissal of the works of Georg Rheticus as rubbish when that same work served as an inspiration for François Viète?

The European understanding and conceptualization of trigonometry in the first half of the 19th century was not so very different than that of ours today. There were six basic trigonometric functions: sine, cosine, tangent, cotangent, secant, and cosecant. These functions could be defined in terms of ratios of sides of a planar right triangle, in terms of a circle of unit radius, or in terms of cartesian coordinates, but neither the right triangle definitions nor the unit circle definitions with which we are so familiar were known to 15th- and 16th-century European mathematicians. The concept of function was not present, nor that of Cartesian coordinates. Angles and triangles were not the central objects of concern, but rather circles, spheres, arcs, and chords. The very words *trigonometry, tangent, secant, cosine, cotangent, and cosecant* were only introduced during the latter half of the 1500’s. Toward the end of this period Viète was only beginning to develop his symbolic algebra, and few if any of his contemporaries understood its power and had the knowledge to use it effectively. Problems were approached through the use of proportionality (the equality of ratios), rather than the solution of equations. Greek mathematical thought about homogeneity was still dominant.[5]

The history of what we now call trigonometry is a long and complex one. With roots firmly in the study of the motions of the stars, sun, moon, and planets, it began in Greece with the study of arcs and chords of a circle on the surface of a plane, or with great and small circles and their arcs and chords on the surface of a sphere. It later spread to India and thence to the Middle East, especially the regions of modern Iran and Iraq. Around 950 AD this knowledge spread to North Africa, and thence to Moorish Spain, and eventually to France, Germany, Italy, and other parts of Europe, where it remained frozen in its medieval form until about 1500. Between 950 and 1500 the eastern centers of Islamic scholarship continued to develop a sophisticated trigonometry, but that knowledge did not reach Spain, Europe, or the Western parts of the Islamic world. For a detailed and documented overview of the development of mathematical astronomy during those years see (Van Brummelen 2009, 2013, 2014).

5.1 The Mystery of the Term "Triplicity"

In defense of Delambre and De Morgan, let me emphasize that both Delambre's histories of astronomy[6] and De Morgan's Penny Cyclopædia were encyclopedic works of thousands of pages, and the amount of time that they could have spent researching small topics such as Torporley's Diclides or Napier's Circular Parts would be very limited indeed; but before we address the different motivations, methods, and conclusions of Torporley and Napier, let us pause to address their common usage of the term *triplicitas*. De Morgan found the use of this term in a mathematical work so unusual that he felt it likely that Napier had read Torporley's work and developed it without giving Torporley credit. In the early modern period, judicial astrology referred to those aspects of astrology that were deemed heretical by the Catholic church in contrast to "natural astrology" such as medical or meteorological astrology.

Napier, in his *Descriptio* bases his analyses on the Astronomical or Navigational Triangle: a spherical triangle formed by the location of the sun, the observer's zenith, and the celestial pole, along with the great circles which are the equators of these three points: the terminator (the projection of the dividing line between day and

[5]Only homogeneous magnitudes are to be compared with one another where comparison means, on the one hand, adding and subtracting magnitudes to form algebraic expressions and, on the other, equating magnitudes or expressions with one another (Klein 1976).

[6]Delambre's six-volume history of astronomy included: Histoire de l'astronomie ancienne (1817), Histoire de l'astronomie du moyen age (1819), Histoire de l'astonomie moderne. 2 volumes (1821), and Histoire de l'astronomie au dix-huitième siècle (1827). The history of astronomy in the 18th century was withheld by Delambre to be published posthumously.

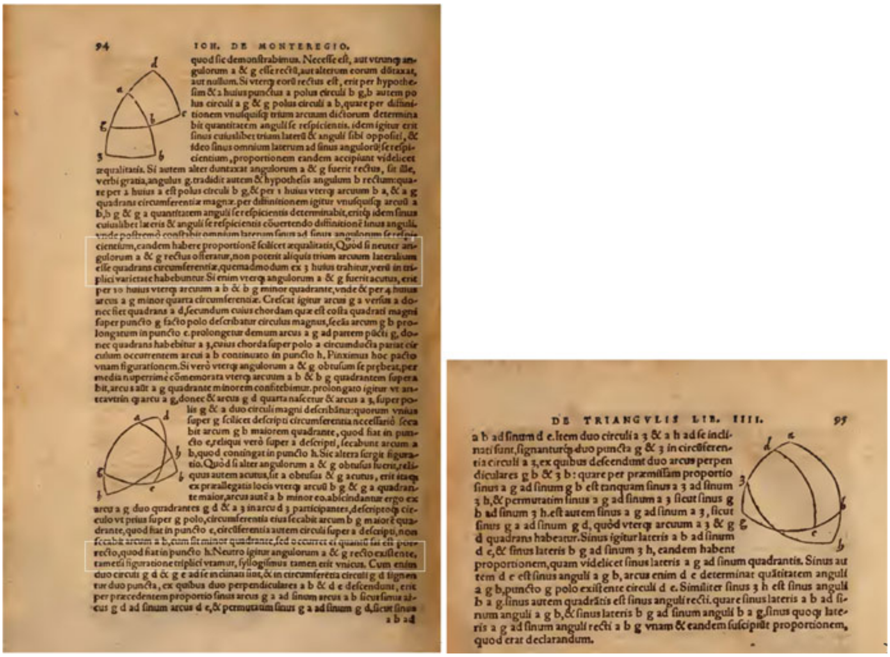


Fig. 1 Book IV, Theorem 16. Regiomontanus : In every right triangle the ratio of the sines of all the sides to the sines of the angles which [the sides] subtend is the same

night on the globe), the observer's horizon, and the celestial equator. To these he adds the solar and observer's meridians. (Napier 1614b) Napier also specifically mentions Regiomontanus as a source for his investigations. Torporley, on the other hand, bases his analyses upon the Hellenistic works of Menelaus and Ptolemy, along with the contributions of Regiomontanus. Since both Napier and Torporley credit Regiomontanus in their publications, exploring Regiomontanus' *De triangulis omnimodis* (Regiomontanus 1533), his masterful summary of all that was known in Europe of the science of triangles at the time of its writing (1464), seemed like a worthwhile endeavor, and in fact the term appears repeatedly in the theorems of Book IV of that work (Figure 1).

If *ABG* is a triangle with right $\angle B$, then ... It is necessary that each of the angles *A* and *G* be a right angle or that only one of them [be a right angle] or that none [of them be a right angle].

- each of the angles *A* and *G* be a right angle
  If each of the angles is a right angle, point A will be the pole of circle BG, B will be the pole of circle AG, and G will be the pole of circle AB. Therefore ... each of the three mentioned arcs will determine the size of the angle opposite, and the sine of the angle opposite that side will be the same, and thus the sines of all the sides to the sines of the angles opposite them will have the same ratio – namely equality.
- only one of the angles *A* and *G* is a right angle
  But if only one of the angles *A* and *G* is right, let that angle be angle *G*, for example.

Moreover, the hypothesis gave angle *B* to be a right angle. Therefore *A* is the pole of circle BG and each of BA and AG is a quadrant. Then by definition each of arcs AB, BG, and GA determine the size of the angle opposite itself and the sine of any side and the sine of the angle opposite will be the same.

- neither angle *A* nor angle *G* is a right angle
  But if neither of the angles *A* and *G* is given to be a right angle, some one of the three arcs comprising the sides cannot be a quadrant; indeed, they may be in a **triplicity** of combinations.[7]

Case (1) If *A* and *G* are both acute, each of arcs *AB* and *BG* will be less than a quarter of the circumference. And each of those arcs can be extended.

Case (2) If each of *A* and *G* are obtuse, each of arcs AB and BG will exceed a quadrant, with arc AG will be less than a quadrant.

Case (3) If one of the angles *A* and *G* is obtuse and the other acute, let *A* be obtuse and *G* acute; then each of the arcs *BG* and *GA* will be greater than a quadrant and arc *AB* will be less than a quadrant. . . .
Thus when neither of the angles *A* and *G* is a right angle, although we use **a triplicity** of constructions, the deductions will nonetheless be one and the same for all three.[8]

Both Napier and Torporley refer, at least in passing, to the works of prior astronomers and mathematicians from whom they drew inspiration. Toporley mentions Regiomontanus, Euclid, and Petrus Ramus; Napier mentions Regiomontanus, Bartolomeo Pitiscus, Thomas Fincke, and Adrianus Metius.

The term "triplicity" appears repeatedly in Regiomontanus' great theoretical work *De triangulis omnimodis,* and both Napier and Torporley refer to Regiomontanus in their own work. It is clear that the term triplicity in Regiomontanus is in no way a technical term of judicial astrology, but rather a mathematical term meaning a trichotomy, or a set of three mutually exclusive cases. Hence we may discard De Morgan's argument that the use of the term *triplicity* is strong evidence that Napier most likely had seen Torporley's work.

6 Torporley's Configurations

Torporley's spherical geometry was firmly based upon the Menelaus figure, a spherical form known at least from the time of Claudius Ptolemy, and perhaps a century before that (Ptolemy et al. 1528; Ptolemy 1998). Astronomers of the time of Torporley would have been well versed in the work of Ptolemy and perhaps with extensions contributed by early Islamic astronomers, and certainly with the work of Regiomontanus (Figures 2, 3 and 4).

[7] Quod si neuter angulorum *a* & *g* rectus offeratur quemadmodum ex huius trahitur, verum in **triplici** varietate habebuntur.

[8] Neutro igitur angulorum *a* & *g* recto existente, tamersi figuratione **triplici** utamur, syllogismus tamen erit unicus.

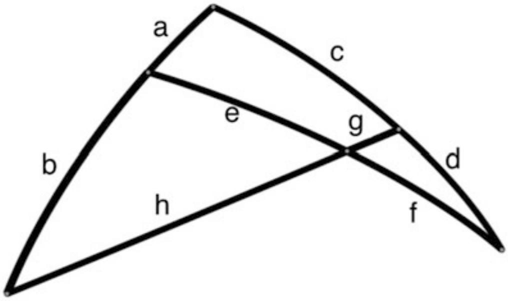


Fig. 2 The Menelaus Configuration

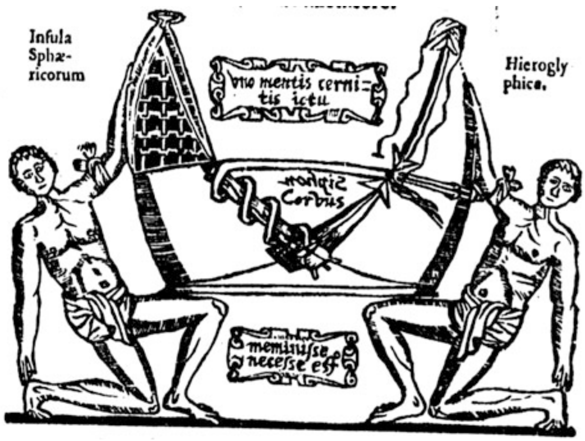


Fig. 3 The two gates combined in mirror image. Note that Corvus and her daughters are viewed from in front of the page, but Siphon and her daughters are viewed from behind the page

6.1 The Spherical Menelaus Theorem

In Hellenistic times this configuration was called a spherical quadrilateral. In medieval Islam, this configuration was called the Sector or the Transversal Figure. In Latin translations theorems were referred to as the *Regula sex quantitatem* or the Rule of Six Quantities.

The disjunctive rule

$$\frac{\sin\ a}{\sin b} = \frac{\sin\ (c+d)}{\sin d}\ \frac{\sin\ g}{\sin\ h}$$

The conjuctive rule

$$\frac{\sin(a+b)}{\sin\ a} = \frac{\sin(g+h)}{\sin\ g}\ \frac{\sin\ f}{\sin(e+f)}$$

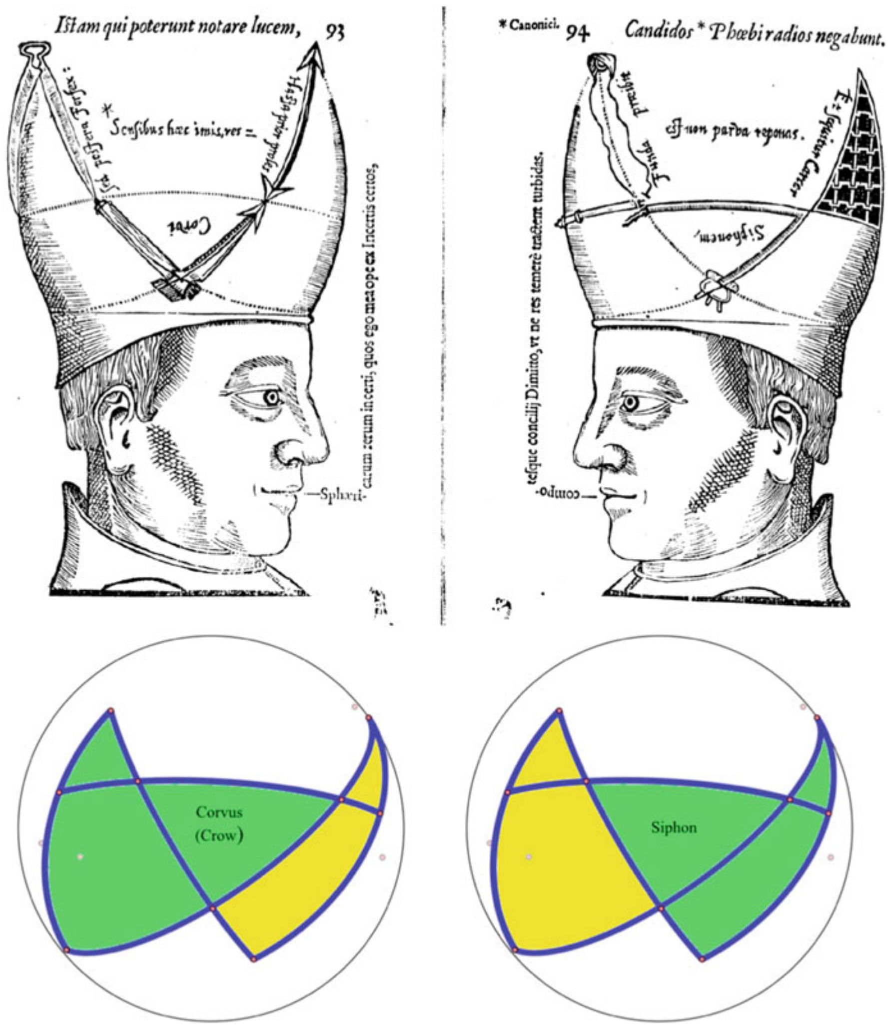


Fig. 4 Each bishop's mitre is composed of two Menelaus figures which overlap in the triangle Crow for the left mitre or the triangle Siphon for the right mitre. Details on the construction of these figures may be found in (Torporlaeo 1603; Silverberg 2009; Delambre 1821)

In the special case where $a+b=c+d=e+f=g+h=90°$, the **disjunctive** case leads to what became known as the Spherical Law of Tangents, embodied in the table Quadrans for the Porta Sinistra

$$\frac{\sin\ a}{\sin(90° - a)} = \frac{R}{\sin\ d}\ \frac{\sin\ g}{\sin(90° - g)}$$

$$\tan\ a = \frac{R}{\sin\ d}\ \tan g$$

or

$$\frac{\sin\ d}{R} = \frac{\tan\ g}{\tan\ a}$$

while the **conjuctive case** leads to the Spherical Law of Sines, embodied in the tabula Quincunx for the Porta Dextra $\frac{R}{\sin\ a} = \frac{R}{\sin\ g}\ \frac{\sin f}{R}$ or more simply,

$$\frac{\sin\ a}{R} = \frac{\sin\ g}{\sin f}$$

where R is the total sine or sinus totus, that is $R = \sin 90°$.

6.2 Summary

Torporley links three right-angled triangles through a pair of overlapping Menelaus Figures. He defines two situations. Let us call the sides adjacent to the right angle "legs" and the side opposite the right angle the "hypotenuse."

- the two sides adjacent to the right angle are given along with one of the angles opposite a leg.
- the hypotenuse and one of the legs are given along with the angle opposite the given leg.

Thus we have a total of six triangles, which Torporley refers to as six triplicities. The central triangle of each triplet contains all the information to convert between the triangle on its left or the triangle on its right. Thus only two tables are needed to cover the six cases.

The six-fold symmetry is an illusion, and as Napier will show the underlying symmetry is five-fold, and not six-fold.

7 Napier's Configurations

Napier's configuration starts not with a Menelaus figure, but with a simple right-angled triangle, as defined below. If we imagine such a triangle with base and perpendicular less than a quadrant, and the hypotenuse also less than a quadrant, we extend the perpendicular until it has a length equal to a quadrant. The end of this segment will be the pole of the base, and we connect this pole with the unattached end of the base.

7.1 *Of Sphaericall Triangles*

- A Sphaericall triangle is either quadrantall or not.
- A quadrantal, is that whose side or angle is equall to a quadrant: whereby we teach, that the knowledge of a quadrantal that is not right angled may as easily be gotten, as if it were right angled.
- A quadrantall triangle, is either manifold, or single.
- A manifold quadrantall, is either three right angled, or two right angled.
- A three right angled triangle, is that, whereof every part is equall to a quadrant.
- Therefore every triangle, each of whose three parts not being opposite, are equall to a quadrant, is three right angled.
- A two right angled triangle, is that whereof two angles onely, and the sides subtending them, are severally equall to a quadrant.
- In every two right angled triangle, the oblique angle is equall to his subtending side.
- Every triangle, whereof any part is equall to a quadrant and any oblique angle, equall to his subtendant, is two right angled.
- Every triangle having any two parts severally equall to a quadrant, and the third unequal, is two right angled.
- All the rest are called single quadrantals. (Napier 1618)

7.2 *Of Single Quadrantals*

- A Single Quadrantall, is that whereof one part onely is equal to a quadrant, and the other five parts are not quadrants.
- Of these five parts which are not quadrants, those three which are furthest removed from the right angle, or the side that is a quadrant, we turne into their complements, and retaining the old order, we bring them all five into a circular or quinquangled situation, and we call them Circulars.
- Hence it is that there bee many triangles, not conformable in their natural parts, which in these Circular parts, doe altogether agree, and are resolved by this our method of Circulars.
- *This uniformitie of the Circular parts, most manifestly appeareth in right angled triangles, made on the superficies of a globe, of five great circles, the first whereof cutteth the second, the second the third, the third the fourth, the fourth the fifth: and lastly, the fifth the first, at right angles. But all the other sections shall be made at oblique angles.*
- The same uniformitie of the circular parts appeareth also in quadrantals that bee not right angled, made upon the superficies of a Globe out of five points, the first whereof is distant from the second, the second from the third, the third from the fourth, the fourth from the fifth, and the fifth from the first by distances and arches equall to a quadrant; but the other distances of the poynts bee unequall to a quadrant.
- Of the five circular parts, three alwayes come in question: whereof the two first are given, the third is sought for. *And of these three, one is in the middle and two are the extreames, which are either set about the middle, or opposite to it. The Logarithme of the middle one is equall to the Differentials of the extremes set about it, or to the Antilogarithmes of the opposite extreames.* (Napier 1618)

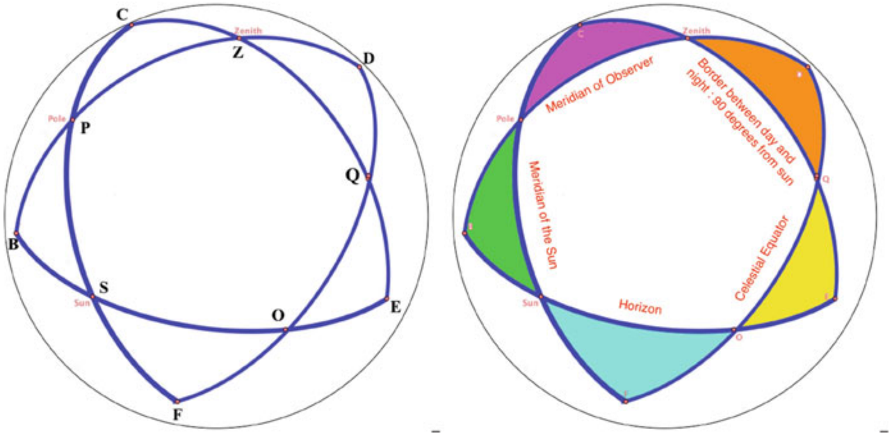


Fig. 5 Five Great Circles creating a cycle of five single quadrantal right-angled triangles, each with identical circular parts, appearing in the same order with differing starting positions, although having differing natural parts

7.3 Napier's Construction of Five Intersecting Great Circles

Let arc BSOE be the horizon, where B is the North point on the horizon, S is the position of the setting sun on the horizon, and arcs BO and SE are quadrantal arcs. Let arc BPZD be the celestial meridian. Hence the vertex angle at B is a right angle. P is the elevated celestial pole, Z is the Zenith, and the arc PD is a quadrantal arc. Point D is both on the equator (being 90 degrees from the pole) and on the meridian. With center P, construct arc DQOF, the celestial equator, which will intersect the meridian at a right angle. Construct the great circle passing through S (the solar position) and P (the celestial pole). This circle will be the meridian of the sun, and extend SP so that SC and PF are quadrants. The meridian of the sun will be perpendicular to the equator at Point S.

This construction forms five simple quadrantal triangles, each of which is right angled, with no other part being a quadrant. See Figure 5. Also as a consequence of this construction, the great circle arc connecting the points S and Z, Z and O, O and P, P and Q, Q and S will form a five pointed star, each leg of which is a quadrant. This forms five simple quadrantal triangles, whose longest side is a quadrant, with no other part being a quadrant. See Figure 6.

7.4 Summary

Napier does not begin with a Menelaus figure (Figure 7), but with a single quadrantal triangle where one angle is a right angle. He then shows how to extend the base

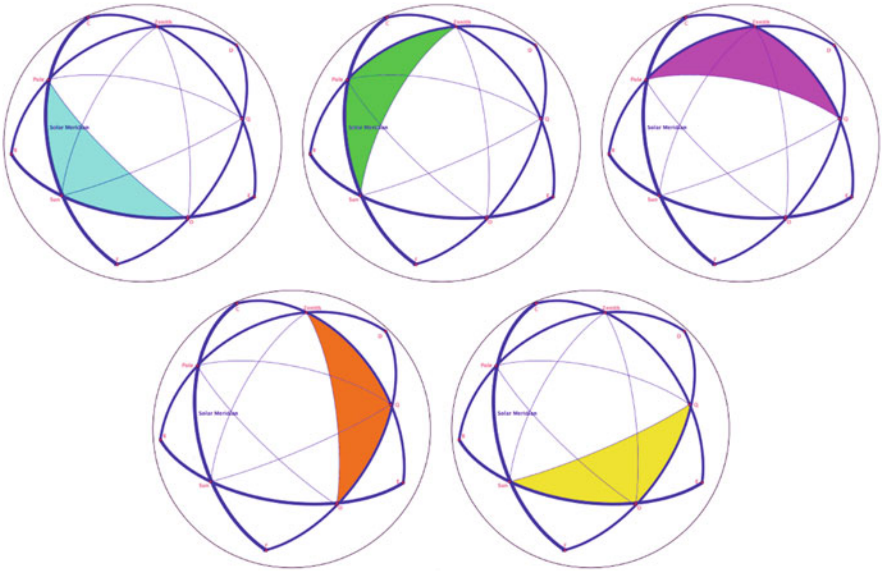


Fig. 6 Five single quadrantal triangles with one quadrantal side, which have identical circular parts, appearing in the same order with differing starting positions, although having differing natural parts

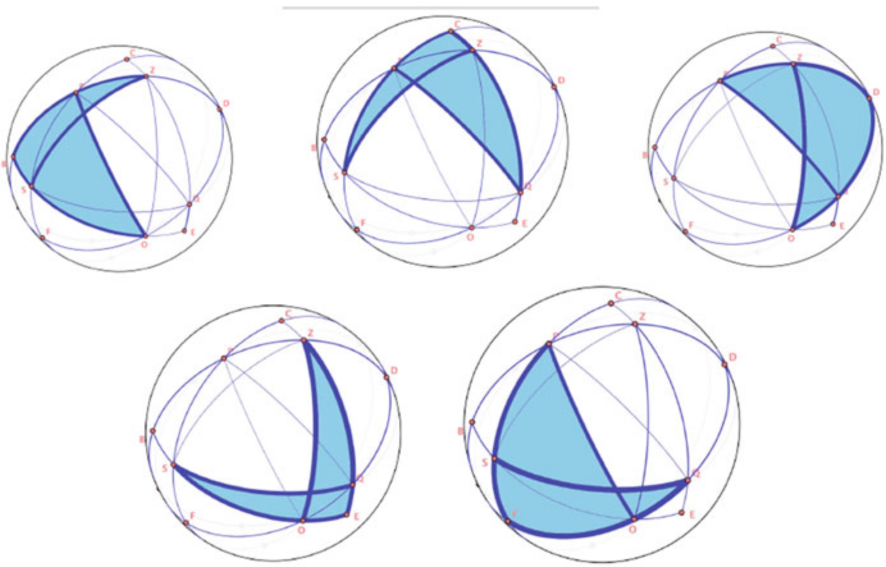


Fig. 7 Although neither the cycle of right-angled triangles, nor the cycle of right-sided triangles displays the structure of a Menelaus configuration, combining both the great circles and the five-pointed star reveal a cycle of five Menelaus configurations. However, Napier does not point this out, nor does he employ the Menelaus figures in developing his Rules of Circular Parts

and perpendicular to form a "complementary" triangle which is a simple quadrantal triangle with one side that is a quadrant. The combined triangles are birectangular (having precisely two right angles); hence, the nonquadrantal side and its opposite angle are equal.

Following this he takes the simple quadrantal triangle with one side that is a quadrant, and adds a suitable right-angled triangle, so that the combined triangles are birectangular as above. Then, comparing the two simple quadrantal right-angled triangles, the circular parts are identical and appear in the same order, but with differing starting positions for each right-angled triangle. Continuing the process a second right-sided simple quadrantal triangle can be formed and compared with the first. Those circular parts will be identical and will appear in the same order, but with a different starting position for each right-angled triangle.

The same procedure can be used, starting with a single quadrantal triangle with a side that is a quadrant. The quadrant can be extended by adjoining a single quadrantal triangle with a right angle. That triangle can be extended to produce a second right-sided triangle. The two right-sided triangles will have identical circular parts, appearing in the same order, with differing starting positions for the cycle.

8 Conclusions

Napier's analysis, however, reveals a five-fold symmetry which is not apparent in Torporley's work.

The circularity at the heart of Napier's Rules of Circular parts is not present in Torporley's approach and has been overlooked by De Morgan and others who have claimed that the Napier's work is essentially a cleaned up version of that of Torporley.

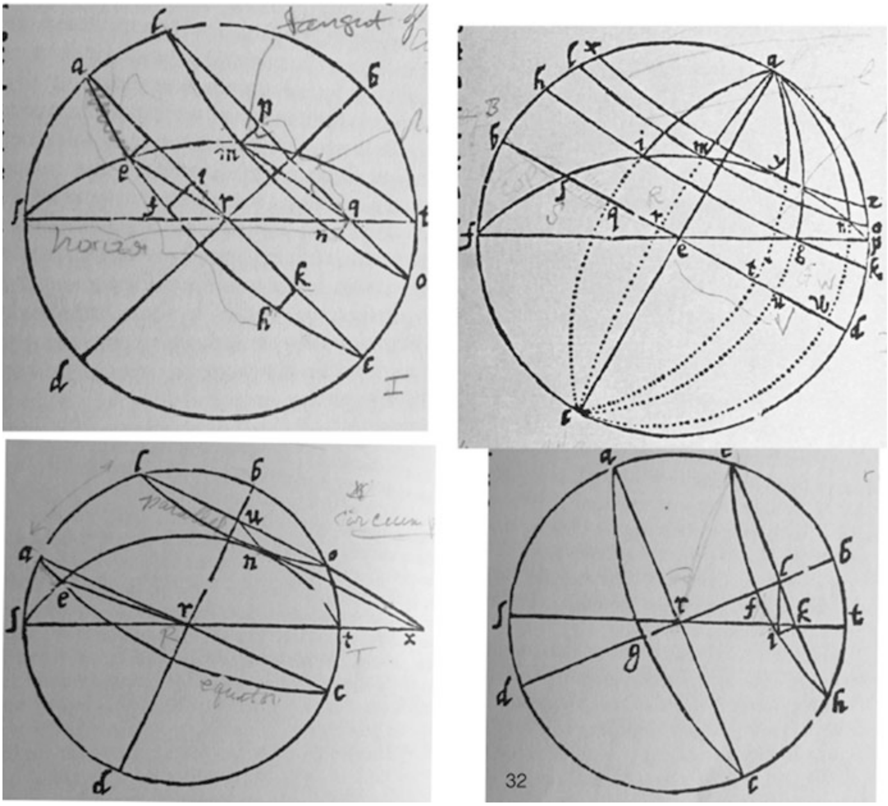
Napier's rules are an insightful look into the structure of rectangular spherical triangles. They illuminate the constraints of their confinement to the surface of the sphere. All of the properties of the self-polar star-pentagon developed by Napier are already present in the very first right triangle drawn.

The cyclical properties of this chain of triangles, and the ways in which the cyclical parts behave differently depending on whether or not they are adjacent to the right angle or the quadrantal side are insights properly attributed to Napier, rather than Torporley.

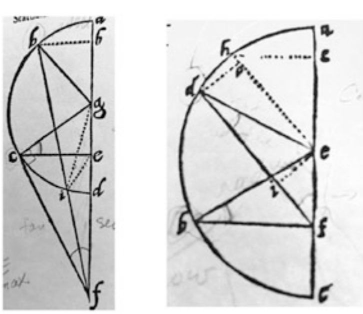
Torporley, on the other hand, was not looking for circularity, symmetries, or repeating patterns. He considered aspects internal to and external to the sphere, as well as those upon its surface, and although spherical trigonometry requires the sides of all triangles to be arcs of great circles, Torporley's astrological interests required him to consider daily motions and positions of stars, lights,[9] and planets that moved along small circles during the course of a day.

[9]Sun and moon.

He dealt extensively with projections onto a variety of planes, in particular those of the ecliptic, equator, horizon, and various meridians, as well as planes that did not pass through the center of the celestial sphere, but defined day circles on its surface.



The table for the right portal (siphon and daughters) is generated indirectly, using some additional geometric constructions for plane triangles within the sphere, lying in various planes. The manner in which this table was composed is intriguing, but my understanding of the five-part table is still very much a work in progress.



Although Torporley's transforming every rectangular triangle to one of two "mother" triangles and then providing a table of precomputed solutions for an unknown part can be used in the same way as Napier's rules, their investigations were independent, different in kind, with differing goals. Neither gentleman's contributions were trivial, although Torporley's were more directly modelled on classical Greek and Islamic works, and Napier's more refined and perhaps more modern; each of them merit credit and deserve to be understood in the context of their times and motivations.

References

Augustus De Morgan. Circular parts (Napier's). *The Penny Cyclopaedia for the Diffusion of Useful Knowledge*, VII:195, 1837.

Augustus De Morgan. On the invention of the circular parts: To the editors of the philosophical magazine and journal. *Philosophical Magazine and Journal*, Series 3, Volume 22(146):350–353, 1843a.

Augustus De Morgan. Nathaniel Torporley. *The Penny Cyclopaedia for the Diffusion of Useful Knowledge*, XXVI:312, 1843b.

Jean Baptiste Joseph Delambre. *Histoire de l'astronomie moderne*. Courcier, 1821.

J. Klein. *Greek Mathematical Thought and the Origin of Algebra*. MIT Press Classics. MIT Press, 1976. ISBN 9780262610223.

John Napier. Mirifici logarithmorum canonis constructio, et eorum ad naturales ipsorum numeros habitudines; vna cum appendice, de alia eaque præstantiore logarithmorum specie condenda. quibus accessere propositiones ad triangula sphaerica faciliore calculo resoluenda: vnà cum annotationibus aliquot doctissimi d. henrici briggi in eas, et memoratam appendicem. authore et inuentore ioanne nepero, .., 1614a.

John Napier. *Mirifici logarithmorum canonis descriptio, ejusque usus, in utraque trigonometria, ut etiam in omni logistica mathematica, amplissimi, facillimi, et expeditissimi explicatio.* ex officina Andrea Hart biblipola, Edinburgi, 1614b.

John Napier. *A Description of the Admirable Table of Logarithms translated in English by Edward Wright. All persued and aproued by the Authour, and published since the death of the Translator.* Simon Waterson, London, 1618.

Claudius Ptolemy. *Almagest [Translated by G.J. Toomer].* Princeton University Press, 1998.

Claudius Ptolemy, Luca Gaurico, and George. *Clavdii Ptolemaei ... Almagestvm Sev Magnae Constrvctionis Mathematicae Opvs Plane Divinvm Latina Donatvm Lingva*. [& calcographica Luceantonii Iuta officina Aere proprio, ac Typis excussa], [In Vrbe Veneta], 1528.

Johannes Regiomontanus. *De triangulis omnimodis libri quinque*. John Petreus, Nurnberg, 1533. URL https://books.google.com/books?id=lVxRAAAAcAAJ.

Joel Silverberg. Napier's Rules of Circular Parts. In *CSHPM, University of British Columbia, Vancouver BC, Canada*, volume 21, pages 160–174. Canadian Society for History and Philosophy of Mathematics, 2008. URL https://www.academia.edu/5870432.

Joel Silverberg. Nathaniel Torporley and his Diclides Coelometricae (1602)–a preliminary investigation. In *CSHPM, Memorial University, St.John's, Newfoundland, Canada*, volume 22, pages 143–154. Canadian Society for History and Philosophy of Mathematics., 2009. URL https://www.academia.edu/5890336. CSHPM, Memorial University, St Johns, Newfoundland, Canada.

Nathale Torporlaeo. *Diclides coelometricae seu Valuae astronomicae vniuersales omnia artis totius munera pseuphophoretica in sat modicis finibus duarum tabularum*. Excudebat Felix Kingston, Londini, 1603.

Glen Van Brummelen. *the Mathematics of the Heavens and the Earth, The early history of trigonometry*. Princeton University Press, 2009.

Glen Van Brummelen. *Heavenly mathematics: The forgotten art of spherical trigonometry*. Princeton University Press, 2013.

Glen Van Brummelen. *Triangles Before Logarithms: Trigonometry in the Lost Century*. 2014. URL https://www.youtube.com/watch?v=-OoyZtwhGbY.

The Reception of American Mathematics Education in Soviet Pedagogical Journals of the 1960s and 1970s

Mariya Boyko

Abstract North American historians of mathematics education have provided detailed accounts of the 1960s "new mathematics" movement, its goals, features and aftermath. Parallel to the reforms in the West, but somewhat later, innovative and fundamental changes to mathematics education were being carried out in the Soviet Union. Soviet educational theorists were aware of the Western developments and discussed them in periodicals devoted to mathematics education. The Soviet reforms and their legacy have not been covered adequately in the literature. The paper will examine the reception of American mathematics education in Soviet pedagogical journals of the 1960s and 1970s and provide a comparison of the Russian experience with what took place in the West.

Keywords New math • Soviet Union • Mathematics education • Journals

1 Introduction

The decades that followed the end of the World War II were marked by political and social turmoil associated with the onset of the Cold War and its consequences. The Union of Soviet Socialist Republics and the United States of America were constantly trying to surpass each other in terms of creating advanced space technologies and weapons of mass destruction. However, military glory and sci-fi-like ideas of space exploration were not the only topics preoccupying the governments and the societies of both countries. Public education was the subject of concern for politicians and the general population alike.

The mathematics-curriculum reform called the "new mathematics" movement, which took place in the USA in the 1950s, and its goals, features and aftermath have been thoroughly analysed by historians of mathematics and education. A decade later, almost parallel to the reforms in the West, innovative and fundamental

M. Boyko (✉)
University of Toronto, 91 Charles Street West, Toronto, ON, Canada M5S 1K7
e-mail: mariya.boyko@mail.utoronto.ca

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*,
Proceedings of the Canadian Society for History and Philosophy of Mathematics
La Société Canadienne d'Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_11

mathematics-education reforms were carried out in the Soviet Union. They were often referred to as "Kolmogorov's reforms", after the prominent Russian mathematician and educator Andrei Kolmogorov. In 1970 he became the head of the committee responsible for rewriting the mathematics curriculum. This committee consisted of professional mathematicians and teachers. However, most of them were either completely immersed in mathematical research or specialized in teaching gifted high school students or university students who were already interested in mathematics. As a result, the reformed curriculum was influenced more by trends in modern mathematical research rather than by pedagogical innovations. High school teachers often criticized the new curriculum for its changes in methods of presentation of the material and insufficient time for adjusting to the new guidelines (Abramov 2010, 81–140). Although teachers and curriculum authors often published their opinions in pedagogical journals, such as *Mathematics in the School*, the reforms and their legacy remain largely undocumented in the historical literature.

Soviet and American mathematics-curriculum reforms may seem similar at first glance. The fact that the "new mathematics" movement started earlier than the Soviet reforms may create an impression that Russian educators were using Western ideas and implementing them in local schools. However, the reasons for the implementation of the reforms, as well as the political and intellectual context in which they were carried out, were very different. The US government interpreted the launch of the Soviet satellite Sputnik in 1957 as a "scientific Pearl Harbor" and as a real threat in the space race. The American mathematics curriculum reform benefited from the increased funding of science and mathematics education programs provided by the government as a part of National Defence Education Act following Sputnik. (Walmsley 2003, 13). The Soviet mathematics curriculum reform, in turn, was part of the larger set of government-initiated education reforms of 1958. The latter reforms were intended to bring the school curriculum closer to students' daily lives and to emphasize the practical aspects of each subject. While Soviet educators were aware of the changes in the Western mathematics curriculum, they were never simply borrowing Western ideas.

In this paper we will examine a prominent Soviet journal for mathematical pedagogy called "Mathematics in the School" and document some of the published opinions, criticisms and discussions that Soviet teachers presented regarding Western mathematics education. These discussions occupied a not insignificant place in Soviet mathematics education in the 1960s and the 1970s. Surprisingly, they have not been studied by historians of mathematics in any detail. This void in the literature can be explained by the secretive atmosphere surrounding the Soviet Union, hidden behind the "Iron Curtain", and by the seeming lack of necessity for providing professionally written historical rather than pedagogical accounts of the reforms.

Many authors have discussed the peculiarities of the teaching techniques and new ways of presenting ideas to students that emerged during the 1960s and the 1970s in Russia.[1] However, the accounts that address the political, social, intellectual and historical contexts of Kolmogorov's reforms have either been brief, or have focused on general aspects of education reform, without emphasizing the changes in the mathematics curriculum in particular. Nevertheless, numerous primary sources such as pedagogical journals, textbooks and articles dealing with mathematics education are now available through digitalized archives of government-issued documents or libraries. Many of these primary sources are in Russian and have not been translated into English. This paper will seek to provide more information on the topic of the Kolmogorov reforms and to give a brief history of education in Russia in order to better understand the historical context in which these reforms took place.[2]

2 Comparison with the "New Mathematics" Movement in America

The Soviet mathematics-curriculum reform of the 1960s did not initially imply any drastic changes to the curriculum's academic content. However, the presentation of the content changed. The new curriculum involved the introduction in elementary school of set theory and a deductive logical approach to the subject. It is important to note that an emphasis on set theory and logical deduction was not novel in the international community of mathematics educators. From the 1950s into the 1960s the USA had undergone a major mathematics-curriculum reform – referred to as the "new mathematics movement", or simply the "new math".

By the 1940s American society had begun to recognize the limits of traditional mathematics education. Youth lacked basic mathematical skills. This was most evident in military settings, where army recruits were often unable to succeed even in activities related to bookkeeping. A program called the "life adjustment movement" emerged in the mid 1940s. The proponents of this movement – mostly members of the education community at the early stages, later joined by the general public – stated that over 60% of students did not possess the intellectual skills that would enable them to go to college or to hold a position requiring specific skills. Hence, they proposed that the new courses in mathematics must focus on purely practical applications of knowledge. Home economics, insurance and taxation were favoured over algebra, geometry and trigonometry – these were excluded completely. The life adjustment movement was resisted and criticized by groups of parents and journalists for dramatically reducing and simplifying the academic content of the mathematics curriculum. However, most educators favoured it and even demanded to increase the level of its availability (Klein 2003).

In the early 1950s American student's knowledge of mathematics continued to decline. There is no evidence to assume this was a direct result of the life-adjustment

[1] Such authors include Alexander Karp, Alexander Abramov and Igor Kostenko.

[2] Unless otherwise noted, the translations from the Russian in the present essay are by the author.

movement. However, it was evident that students' performance was not improving. At the same time, the political tensions with the USSR, and the arms race and the space race led to a demand for a steady flow of specialists qualified in mathematics and physics. As a result, the government granted substantial funding to mathematics education (Walmsley 2003).

A new group of progressive educators, who generally stood for hands-on discovery-based learning which would help students develop social responsibility and critical thinking skills, concluded that "mathematical education had failed because the traditional curriculum offered antiquated mathematics, by which they meant mathematics created before 1700" (Kline 1973, 17). These educators did not appreciate the fact that mathematics is a cumulative discipline and that students need to firmly grasp the older concepts before they can proceed to learn and understand modern research. Nevertheless, the idea of modernizing the curriculum took hold in the community of educators and the new mathematics movement was implemented throughout the 1950s.

New math was marked by an emphasis on formal notation, concepts of set theory, the structural laws of algebra and an axiomatic approach to the subject, starting at the elementary school level. For instance, elementary and secondary school students were expected to understand the distributive, associative and commutative laws of algebra even if their mathematical skills were not yet strong. The new math was developed and implemented in a different social and intellectual context from Kolmogorov's reforms and was in decline by the 1970s. Nevertheless, the characteristic criticisms of the new mathematics movement would also apply to the reforms advocated and implemented in the Soviet Union by Kolmogorov and his proponents.

During the nineteenth century mathematics expanded greatly, with whole new subjects being invented and existing subjects being expanded deepened and transformed. As the century came to a close mathematicians increasingly used a deductive-logical approach to formulate new results and subject areas. Hilbert (1990) emphasized the importance of rigour and clarity as well as an axiomatic approach to various mathematical subjects, of which the theory of probability was an important example. Prominent mathematicians in the first half of the twentieth century such as Emmy Noether, Bartel van der Waerden, the Bourbaki group, and Saunders MacLane believed that an increased emphasis on rigour, a focus on the concept of mathematical structure and an axiomatic approach were the signal characteristics of "modern" mathematics. (See Corry (1997) for an account of the emergence of modern mathematics.) Kolmogorov was well known for his work in probability in the 1930s, and the formal deductive approach he followed – including his famous axioms – in presenting his results.

The educational reform movements in mathematics in the second half of the twentieth century were a response to a complex variety of interests. Among these were professional and governmental organizations, as well as changes in demographics, industrialization, urbanization, and so on. At a more general level changes in the ideology of research mathematics and the emergence of what is known as "modern" mathematics, influenced the thinking of Western and Soviet educators alike, and must be taken into account in understanding the wider theoretical culture that informed educational reform. Among other things, the authority of Kolmogorov – in many respects a stereotypical modern mathematician – in the world of professional and educational mathematics was very substantial in both the USSR and in the USA.

Some professional mathematicians in America believed the introduction of set theoretic notions would only create confusion among elementary school students, especially because the "material about sets [was] never used – nor [was] any explanation given as to why the concept is of any particular interest or utility." The physicist Richard Feynman asserted that "often the total number of facts that are learned [was] quite small, while the total number of words [was] very great" (Kline 1973, 69).

It is important to note that American and Russian schools were often understaffed in the 1950s, 1960s and 1970s. Moreover, teachers also lacked time to learn the ins and outs of the new curriculum. Insufficient preparation of school teachers and the lack of an adequate time frame to prepare the new curriculum documents contributed to the ineffectiveness of the reforms in both countries (Wu 1998). The principal authors of the new curriculum in the USA and USSR were teachers and professional mathematicians who were accustomed to working with gifted students. However, a curriculum that was designed for gifted students could not be implemented for a general audience without a thorough training of teachers and modifications of existing pedagogical techniques. Unfortunately, the time frame of the reforms in both the USA and the Soviet Union did not allow for a re-training period. Even though additional courses were offered to teachers, educators could rarely complete them due to their heavy workload (Walmsley 2003).

3 The Reception of American Mathematics Education in the Soviet Union

American mathematics educators launched the new math curriculum in the early 1950s and had identified its shortcomings by the early 1970s. Soviet mathematics reforms began in the late 1950s and continued for the next three decades, finally falling out of favour with the Soviet regime by the 1980s. Given the parallel nature of the two reforms, is it natural to assume that the programs borrowed ideas from one another. More precisely, did the proponents of Kolmogorov reforms borrow from the new math?

There are many reasons to conclude that direct borrowing was unlikely. Listing and discussing all of them is beyond the scope of this article. We will focus on the reflections and impressions of Western mathematics education in Soviet pedagogical journals. It will become evident that Soviet educators were informed about the American trends in mathematics education. However, they were not inclined to implement these trends directly in the Soviet setting. Western innovations in the teaching of mathematics were often presented in a neutral light without substantial praise or criticism. Even the rare cases of harsh criticism of Western educators were rather a product of biases which were specific to the particular author.

The pedagogical journal *Mathematics in the School* has been publishing various articles related to teaching mathematics abroad since the 1930s. However, the frequency of the appearance of such articles and the countries that they were dedicated to varied over different time periods. For instance, there were numerous articles dedicated to teaching of mathematics in America and Europe, as well as Asia, during the 1930s and 1940s. Many of them were neutral in terms of political and ideological content. Articles dedicated to pedagogy in non-Communist foreign countries during the 1940s emphasized the superiority of the Communist state over capitalist states. Professor of mathematics Ivan Depman wrote in his 1949 article "Some Information about the State of Mathematics Education in Contemporary Foreign Schools" (Depman 1949, 39) that the mathematics curriculum in capitalist countries, including the USA, relied too heavily on textbooks that focused purely on "arithmetic for commerce." He added that these textbooks lacked ideological education, and all of the proposed tasks and word problems were intended to teach "commercial transactions" (Depman 1949, 39). He then cited other Soviet authors who were familiar with American mathematics education and concluded that the majority of American youth lacked an adequate knowledge of basic mathematics. According to Depman, students in the USSR were able to master the tasks from the American mathematics tests much better than their American counterparts. However, it was not clear which specific groups of Russian students Depman referred to, since he only mentioned that he asked several teachers from Leningrad to give their students some mathematics tests produced in the USA. The teachers had proposed these tests to the students when a free hour of class time was available (Depman 1949, 40). Although this evidence is significant, the results cannot be considered reflective of the situation in the entire USSR. Depman asserted that the mathematics that the American students were learning in high school was insufficient for entering post-secondary education programs in engineering and science-related fields (Depman 1949, 48). Although his and other articles were clearly biased in favour of Soviet education, the Soviet educators and American educators were at least in agreement that the mathematical level of American students was insufficient and needed improvement.

Over the course of the 1950s when the Cold War was at its peak, articles on foreign teaching methods published in *Mathematics in the School* focused primarily on European countries and Communist countries. Policies and pedagogical techniques of communist countries were represented more favourably than the experiences of teachers from capitalist countries. For instance, the teaching of mathematics in France was described as too theoretical (Depman 1949, 40). The article "On Teaching of Trigonometry in Some Foreign High Schools", published in 1957, indicated that the US curriculum lacked coherence in comparison with the USSR curriculum (Lebedev 1957).

Articles on American mathematics education appeared regularly in *Mathematics in the School* during the 1960s, although the term "new mathematics" was not used to characterize the new tendencies in education. Published articles demonstrate that Soviet educators, professional mathematicians and authors of the reformed Soviet curriculum were aware of Western developments in teaching mathematics. These articles were informative in nature and did not carry political meaning. For instance, an article by D. I. Marchenko, "Overview of Algebra Textbooks for Public Schools of the U.S.A.", published in 1961, summarized the American mathematics education system and presented excerpts from textbooks used by Western educators. Marchenko stated that American mathematics teachers were free to choose any textbook that suited the intellectual needs of their students. The topics that were included in textbooks were chosen by the authors. Hence, the content of textbooks intended for the same grade level was often very different. The teachers had to be mindful of the peculiarities of each textbook in order to cover all the topics that were contained in the curriculum. Marchenko also presented several excerpts from textbooks that contained word problems and their solutions. The word problems were designed to stimulate logical thinking and to illustrate practical applications of mathematics in everyday life. For instance, Marchenko stated that many word problems were based on a scenario where a swimming pool was being filled with water. Solutions to these word problems did not seem to emphasize rigorous notions of set theory (Marchenko 1961).

In 1962 a direct excerpt from the *Report of the Commission on Mathematics. Appendixes. College Entrance Examination Board* that was originally published in New York in 1959, and translated by the prominent Russian mathematician, educator and historian of mathematics Aleksei Markushevich, was published in *Mathematics in the School*. The very fact of this publication demonstrated that professional mathematicians and leading figures in mathematics education were interested in learning about mathematics educational practices in the West and wished to make this knowledge accessible to teachers across the USSR. The excerpt from the aforementioned book contained instructions for teachers on presenting the topic of irrational numbers to students. Teachers were advised to specify the difference between rational and irrational numbers to the students and then explain the nature of irrational numbers in multiple ways. For instance, it was suggested that a diagram be presented to students to assist a numerical understanding of an irrational number American Committee for Mathematical Pedagogy 1962).

Articles on American mathematics pedagogy that appeared in *Mathematics in the School* in 1964 and later in the decade contained more information on the ways that Western educators were using set theory to teach mathematics to students of school age. V. B. Yudina summarized information that was originally published in the American journal *The Mathematics Teacher*. She called attention to the importance that Western educators placed on the teaching of symbolic logic and basic concepts of set theory. For instance, students were expected to be familiar with concepts of union and intersection of sets and to construct truth tables. She reported that the numerous diagrams and charts that were used to aid the understanding of these concepts could be challenging for the students. Yudina also emphasized that American educators placed importance on the deductive nature of learning and on the teaching of concepts related to symbolic logic (Yudina 1964). 1 The

American author and professor of mathematics education Bruce Vogeli, whose article “Modernization of Teaching of Mathematics in American Schools” was published in *Mathematics in the School* in 1964, summarized the main goals of the mathematics curriculum reforms that were taking place in the U.S.A. He did not use the term “new mathematics” explicitly, but noted that changes that took place in mathematics education over the previous decade were “revolutionary”. Vogeli acknowledged that American educators were dissatisfied with the way students memorized by rote various mathematical facts. He stated that the newly formed American curriculum was designed to encourage the students to make discoveries rather than to simply memorize facts. Students were expected to learn the basic notions of set theory at early stages of elementary school. The concept of negative numbers was also included in the curriculum of elementary school. It was also implied that teachers’ work was being valued and respected, and that students were keeping up with the fundamental requirements of the curriculum when sufficient assistance was provided. Overall, Vogeli presented a positive picture of the new trends in American mathematics education (Vogeli 1964, 88–90). However, his ideas were never explicitly praised or rejected by authors in any of the following issues of *Mathematics in the School*. There is no evidence that Soviet educators were planning to implement any of the American innovations in teaching mathematics, even if they had a beneficial impact on American students.

Another well-known Russian mathematician, educator, and author of mathematics textbooks, Isaak Yaglom, published an overview of various American textbooks on geometry in his article “Geometry in the Schools of the U.S.A.,” published in *Mathematics in the School* in 1967. Yaglom observed that American authors valued an axiomatic approach in learning and teaching various concepts in geometry. He stated that textbooks contain numerous definitions that the students are expected to learn in order to derive more complex theorems and statements. Although he did not give any detailed examples of the ways in which geometry was presented to students in these textbooks, he expressed the concern that such trends in pedagogy might not be suitable in the Soviet setting. Nevertheless, he emphasized that it was important for Soviet educators to be fully aware of the work of their American colleagues (Yaglom 1967, 96).

In the same year 1964 another article on American mathematics education was published by one P. A. Alexandrova in *Mathematics in the School*. Alexandrova examined the American presentation of concepts of arithmetic and algebra in textbooks for students in grades seven to nine. She briefly summarized the curriculum requirements of each grade level and then provided several examples of problems and concepts that American students were expected to learn. Students needed to be familiar with rational numbers, irrational numbers, and integers. They were to have a basic knowledge of set theory which included knowing the properties of various sets, understanding the concepts of union and intersection of sets, and so on. She noted that the concept of function was not fully explained to American students at the early stages of their mathematics education. At first the students were encouraged to develop an intuitive understanding of functions. Later they needed to learn that a function was a relationship between the elements of various sets. A detailed definition of a function and examples of complex functions were presented to the American students at the senior school level (Alexandrova 1967).

One paper that touched upon the new trends in American mathematics education was published by Rolf Nevanlinna in 1968 in *Mathematics in the School* in 1968. Nevanlinna noted that axiomatic and deductive approaches found their way into most areas of modern mathematics. He emphasized that influential mathematicians such as David Hilbert and the Bourbaki group were promoters of this trend. Naturally, these ideas were bound to influence the field of mathematics education. While Nevalinna did not provide detailed comments on the effectiveness of the new methods used by Western teachers, his article was informative and concise (Nevanlinna 1968).

4 Conclusion

It is evident that Soviet educators were aware of the work of their Western colleagues. However, this awareness does not seem to have involved any direct borrowing of Western ideas or any overt influence of American mathematics educators on Russian educators. Even if such influence existed, it can at best be described as indirect. However, modern mathematics was a significant factor that simultaneously influenced American, European and Russian mathematicians and educators. Modern mathematics encouraged an interest in set theory and structures within mathematics. American educators incorporated these ideas into the mathematical school curriculum in explicit ways (such as emphasizing the structural laws of algebra at elementary school level (Kline 1973)). In contrast, Russian educators rarely made explicit references to structures in mathematics at the school level. In both countries the curriculum was rewritten by professional mathematicians and educators who were more accustomed to working with talented children, or even university students (Abramov 2010, 87–140). Their expectations for the overall academic aptitude of ordinary students might have been too high to begin with. Since many authors of the new curriculum in both countries were professional mathematicians it was natural that their ideas on education would have been influenced by the latest trends in mathematical research (Kline 1973). We can conclude that modern mathematics influenced the minds of the curriculum reformers in both countries, while there is relatively little evidence of a direct influence of the new math on Kolmogorov's reforms. Given the political tensions between the USSR and the USA in the 1950s and 1960s, the Soviet educators would have been hesitant to publicly admit any direct borrowing of Western ideas. Doing so could have been viewed as support for Western culture, something that was discouraged during the Cold War period. While further investigation of this point is necessary, the published and unpublished sources that could shed light on it appear to be scarce.

Acknowledgements I thank Craig Fraser for comments on drafts of this paper. I am grateful to Noah Stemeroff for help with the LateX format. I am also grateful to the two referees for their comments on the paper.

References

Abramov, Alexander. (2010). "Toward a History of Mathematics Education Reform in Soviet Schools (1960's–1980's)," in Alexander Karp and Bruce R. Vogeli (Eds.), *Russian Mathematics Education: History and World Significance*, pp, 87–140. New Jersey: World Scientific Publishing Company.

Alexandrova, P. A. (1967). One of the Experimental Programs in Mathematics for Middle Schools in of the U.S.A. Математика в Школе (*Mathematics in the School*), 3. 91–95.

Corry, Leo. (1997). The Origins of Eternal Truth in Modern Mathematics: Hilbert to Bourbaki and Beyond. *Science in Context*, 10. 253–296.

Depman, Ivan. (1949). Некоторые сведения о преподавании математики в современной зарубежной школе (Some Information About the State of Mathematics Education In Contemporary Foreign Schools). Математика в Школе (*Mathematics in the School*), 1. Moscow: Poligrafkniga (Polygraph Book). Archive at http://www.mathedu.ru/journals-collections/

Hilbert, David. (1990). Mathematical Problems. Lecture Delivered Before the International Congress of Mathematicians at Paris in 1900. *Bulletin of the American Mathematical Society* 8 (1902), pp. 437–479. German original published in *Göttinger Nachrichten*, 1900, pp. 253–297.

Klein, David. (2003). A Brief History of American K-12 Mathematics Education in the 20th Century. Pre-print. Final version published in Mathematical Cognition edited by James Royer, 2003. California State University North Ridge. http://www.csun.edu/~vcmth00m/AHistory.html

Kline, Morris. (1973). *Why Johnny Can't Add*. New York: St. Martin's Press.

Lebedev, V.P. (1957). О преподавании тригонометрии в некоторых зарубежных средних школах (On Teaching of Trigonometry in Some Foreign High Schools). Математика в Школе (*Mathematics in the School*), 2. Moscow: Полиграфкнига (Polygraph Book). Archive at http://www.mathedu.ru/journals-collections/

Marchenko, D.I. (1961). Обзор учебников для общеобразователь ной средней школы США (Overview of Algebra Textbooks for Public Schools of the USA). Математика в Школе (*Mathematics in the School*), 3. 63–72.

Nevanlinna, Rolf. (1968). Реформа в преподавании математики (A Reform in Mathematics Education). Математика в Школе (*Mathematics in the School*), 1. 83–89.

Walmsley, Angela L. E. (2003). *History of the "New Mathematics" Movement and Its Relationship With Current Mathematical Reform*. Maryland: University Press of America Inc.

Vogeli, Bruce. (1964). Модэ рнизация преподавания математики в американской школе (Modernization of teaching of Mathematics in American Schools). Математика в Школе (*Mathematics in the School*), 4. 88–90.

Wu, Hung-Hsi. (1998). The Mathematics Education Reform: What It Is and Why Should You Care? University of California. Accessed December 18, 2012. http://math.berkeley.edu/~wu/reform3.pdf

Yaglom, Isaak. (1967). Хеометрия в школах США (Geometry in the Schools of the USA). Математика в Школе (*Mathematics in the School*), 2. 93–96.

Yudina, I. B. (1964). Вопросы теории множеств И математической логики в журнале The Mathematics For Teachers (Topics in the Theory of Sets and Mathematical Logic in the Journal *The Mathematics Teacher*). Математика в Школе (*Mathematics in the School*), 1. 88–92.

American Committee for Mathematical Pedagogy. (1962). Excerpts from Methodological Instructions of the American Committee for Mathematical Pedagogy. Из методических указаний американской комиссии по преподаванию математкики from Report of the Commission on Mathematics. Appendices. College of Entrance Examination Board. New York, 1959. Translated by G.A. Markushevich. Математика в Школе (*Mathematics in the School*), 5. 79–81.

Mathematics in Library Subject Classification Systems

Craig Fraser

Abstract Insofar as library science is concerned, modern classification of mathematical subjects occurred within the larger framework of library classification, a vast project receiving sustained attention in the period from 1870 to 1920. The work of the library cataloguers was carried out against the background of a broad nineteenth-century interest in the classification of knowledge. We explore different views during this period concerning the position of mathematics in the overall scheme of knowledge, the scope of mathematics, and the internal organization of the different parts of mathematics. We examine how mathematical books were classified, from the most general level down to the level of particular subject areas in analysis. The focus is on the Library of Congress classification system in its various iterations from 1905 to the present.

Keywords Library classification • Mathematical books • Analysis • Library of congress

1 Introduction

> The classification of mathematical studies is involved in extraordinary difficulties, and so is the classifying of many mathematical books. The relations of the branches are so intricate, so plastic, so recondite, that it is well-nigh impossible to define them or to comprehend them. - Bliss (1935, 20)

Insofar as library science is concerned, classification of mathematical subjects occurred within the larger framework of library classification, a vast project which drew sustained attention between 1870 and 1920. The two American giants in library work in the formative period of classification were Melvil Dewey and Charles Cutter. In 1876 Dewey published the famous Dewey decimal system of classification, while Cutter's expansive scheme of 1885 would provide the basis

C. Fraser (✉)
University of Toronto, 91 Charles Street West, Toronto, ON, Canada M5S 1K7
e-mail: craig.fraser@utoronto.ca

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*, Proceedings of the Canadian Society for History and Philosophy of Mathematics
La Société Canadienne d'Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_12

for the Library of Congress system. The latter was established in 1905 by James Hanson and Charles Martel, both European immigrants to the United States. In the early twentieth-century additional classification schemes appeared. Among the more notable of these were the subject "system" formulated by the Englishman James Duff Brown and the "bibliographic" system invented by City College of New York librarian Henry E. Bliss.

The practical goal of all classification was information retrieval, allowing, for example, a user to go to a large library, consult the catalogue, and retrieve a given book of interest. The call number given to a book had to be abstract—it could make no reference to any particular library or to the physical arrangement of the books in any library it happened to belong to.[1] The motivation for classification schemes was the appearance of an increasing number of comprehensive libraries with substantial holdings of books from many subject areas. Library science, as the discipline of book classification and cataloguing had come to be known by the 1920s, was not a fundamental science, but an organized and systematic endeavor with the goal of identifying and retrieving information. Of course, librarians were not averse to finding philosophical meaning or justification for the schema they employed, and they appealed to principles of the organization of knowledge, but the final goal was always a practical one. Classification pioneers such as Dewey and Cutter were generalists who were not primarily concerned with any particular subject area. To the extent that their interests were focused on particular fields, these were found in humanistic and social subjects, not in the natural sciences.

Among all of the major systems of book classification, the Library of Congress scheme (LC) was the one that achieved dominance in university and research libraries. In 1870 the US Copyright Office was by legislation placed in the Library of Congress, and the Library received copies of all publications submitted for copyright. The holdings of the Library increased and became more complete than any elsewhere, including the collections of major university libraries and large public libraries. This was particularly true for the small-edition research books and monographs in academic fields. In 1901 the Library established its catalogue card distribution service, which allowed libraries throughout America to receive catalogue cards for all published books. Erdlund (1976, 398), an historian of the Library, comments on the significance of this development: "By including the card distribution service in its functions, the Library, at that time a reference library to Congress with a small constituency consisting almost exclusively of congressmen and their staff members, was adopting a potentially enormous constituency—that of the total American library community." The importance of the LC system in the world of classification was apparent in the years following its establishment. While the major university libraries with their specialized collections containing

[1]LaMontagne (1961, 208) observes that Charles Cutter "foresaw the continuing growth of the library and knew that each change in the shelving of books entailed the changing of 'shelf marks'—a long and expensive process. Cutter therefore decided to abandon fixed location 'and to adopt a method which would allow books to be moved without changing the marks on the catalogues.'" The part in quotation marks is from (Cutter 1882, 6). Cutter's "bench marks" are what we refer to today as call numbers.

many older and foreign-language books continued to maintain a patchwork of local classification systems, the LC has made steady headway up to the present as the dominant and most widely used system of library classification.

The work of the library cataloguers in the decades around 1900 was carried out against the background of a broad nineteenth-century interest in the classification of knowledge. In the next four sections we examine how mathematical subjects were classified, from the most general level down to the specific level of particular subject areas in analysis.

2 Place of Mathematics in Classification Schemes

Prior to, and concurrent with, the development of library classification systems in the nineteenth century there was a great deal of interest in the general problem of the classification of knowledge.[2] Although this problem was a venerable one, going back to the Greeks, it was of special concern in the nineteenth century and was the object of extensive research and discussion of a kind particular to this period. Insofar as the sciences were concerned mathematics occupied a privileged place in this philosophical project. In his recent book "Why Is There Philosophy of Mathematics at All?" Hacking (2014) explores the prominent place that mathematics has played throughout history in the writings of the great philosophers. In a review of this book Siegel (2016, 253) observes that "mathematics has perennially fascinated philosophers, to the point that the philosophy of mathematics is not the philosophy of a special science—like the philosophy of physics or biology—but rather a central field of analytic philosophy." In all of the major classification schemes, mathematics is either placed first in a category of general or fundamental science—along with philosophy; or it is put at the beginning of the natural sciences, followed by subjects that follow a presumed reductionist foundational order that exists among them: physics, then chemistry and finally biology.

Although the majority of the thinkers at the end of the nineteenth century interested in the organization of knowledge were neither scientists nor mathematicians, they possessed definite ideas about the classification of scientific subjects. A recurring theme was the coupling of mathematics with philosophy and its separation from subjects in natural science such as physics, astronomy, and chemistry. This point of view seems to have reflected the general and fairly widespread influence of philosophical logicism on contemporary scientific thought.

The Harvard psychologist Hugo Münsterberg was responsible for the scientific plan of The International Congress of Arts and Science held in 1904 at the St. Louis Exposition. Knowledge was divided into seven divisions, and each division was composed of a collection of departments. The divisions were: (A) normative sciences; (B) historical sciences; (C) physical science; (D) mental science; (E) utilitarian science; (F) social regulation; and (G) social culture (Bliss 1929, 375–380).

[2]For a general survey of work on classification in the nineteenth century see Dolby (1979).

Philosophy and mathematics comprised division A, while physics and the other traditional natural sciences made up division C.

The philosophical view of the place of mathematics within knowledge never really found favor with librarians who worked on the concrete project of book classification. However the affinity of mathematics with logic—if not philosophy—and its separation from the natural sciences was a prominent feature of Brown's 1906 subject classification scheme. Brown held that the classification of books should reflect the classification of knowledge, so that library classification was never a purely contingent project of information retrieval. Brown asserted that logic and mathematics should be grouped together under "Generalia" and should precede all other branches of knowledge, being preliminary to any field of investigation, from physics to economics to philosophy and history, or anything else. Classes of knowledge were given by Brown in the following order: generalia, physical science, biological science, medical science, agriculture and domestic arts, philosophy and religion, social and political science, language and literature, literary forms, and history and geography.

Brown's classification was also distinctive in its positioning of practical subjects adjacent to their presumed theoretical counterparts. Following the section on mathematics there would be books on painting and sculpture. There were two reasons for such an arrangement. First, fine and graphic art exemplified the visual point of view of geometry and could be regarded in some sense as the embodiment of geometric ideas. Second, Brown believed that the subject of visual representation was fundamental in character and that familiarity with it was necessary for its use in various applied fields of investigation and work, hence its placement under generalia. (It may seem odd that the section of the library stacks devoted to mathematics would be followed by books on Flemish art, while books on physics would be located a few aisles over—but such was Brown's rather idiosyncratic notion of book classification.)

A librarian influenced by Brown was Henry Bliss of the City College of New York. More than Brown, Bliss emphasized the affinity of mathematics with philosophy and logic and its separation from science. He wrote two books on the organization of knowledge: the first (1929) was a general and somewhat philosophical work, while the second (1933) was directed more specifically to the classification of books. Bliss regarded mathematics more as a method than a branch of science, and he believed that grounding in logic was proper preparation for its study. He observed (1929, 258) that "Logic is usually regarded as a branch of philosophy and the close relation of philosophical thought to mathematical thought is often affirmed." He had many criticisms of the Library of Congress classification scheme, including the position in LC of mathematics: "In a broader aspect the separation of Sciences in Q from philosophy in B involves such an unscientific and unphilosophic consequence as separating Philosophy of Science from Philosophy of Knowledge, and more generally separating Logic (BC) from Mathematics (QA), despite the claims of both logicians and mathematicians that their studies are inseparable."

The conception of Münsterberg and Bliss was implicitly rejected by thinkers concerned only with the organization and classification of subjects in the natural sciences. For them, mathematics clearly had to be included as a subject area, usually at the beginning of the classification, while philosophy did not appear at all. Any viable classification scheme would need to reflect the place of subjects in the real world. Classifying mathematics with philosophy and separating it from physics and engineering might be sensible in the domain of humanistic thought, but it made little sense in actual practice.

In the Cutter, Dewey and LC classification schemes mathematics is separated from philosophy and grouped with the natural sciences. In the Dewey system, philosophy is placed near the beginning under 100 and is followed by theology (200), sociology (300), and philology (400). The natural sciences comprise the 500s, with mathematics (510) as the first science subject proper, followed by astronomy (520), physics (530), and chemistry (540).

Cutter also grouped philosophy near the beginning under the letter B, where it was followed by religion and theology (C and D), biography (E), history and geography (F), social sciences (H), and natural sciences and applications (L). The presentation of subjects under L followed the order mathematics, physics, chemistry, and astronomy, with designations via subscripts: mathematics (L_B), physics (L_H), chemistry (L_O) and astronomy (L_R). Cutter classified the remaining science categories under the letters M through S: natural history as well as geology and biology (M); botany (N), zoology (O), anthropology and ethnology (P), and medicine (Q).

The LC system seems to have been patterned after Cutter (1891–93) and the placement of philosophy with respect to the natural sciences follows this earlier system. Sayers (1915–1916, 135) observes that "The outline of the [LC] classification is almost directly based upon The Expansive system, as a comparative paradigm of the two will demonstrate." The LC will be the subject of more detailed study in sections 4 and 5.

3 Scope of Mathematics in Classification Schemes

Until the nineteenth-century mathematics was interpreted broadly to include subjects that today would be regarded as part of astronomy, physics, or engineering. But by the second half of the nineteenth century, when library classification systems were being developed, the scope of mathematics had narrowed substantially.

Writers such as Münsterberg and Bliss who viewed logic and mathematics as kindred subjects and grouped mathematics with philosophy adhered to a conception of mathematics that certainly did not include subjects in physics such as mechanics. However, these thinkers did not represent scientists and mathematicians themselves. Among the latter mathematics was a subject that involved traditional logic only very peripherally if at all. Discussions of the scope and relative position of purely scientific subjects in the nineteenth century focused on what was called the hierarchy of sciences, a notion introduced by Comte in 1830 in the second lesson of his *Cours de philosophie positive*. Comte believed that there is a natural progression of scientific subjects, beginning with mathematics, passing through astronomy, physics, chemistry, and biology, and ending with sociology. This hierarchy could be justified on methodological or philosophical grounds, and was often taken for granted for practical reasons. The Comtean hierarchy of sciences was accepted by virtually all of the systems of book classification, and survives to the present.

Among those writers who were primarily interested in the natural sciences, mathematics was placed within science, at the very beginning of the Comtean hierarchy. An important figure was the French physicist André-Marie Ampère who along with Comte and some other French figures of the period was a mathematical empiricist in orientation. These authors separated mathematics completely from philosophy, which tended to occupy a lower position in the overall scheme of knowledge and learning than it had traditionally held. Mechanics was a kind of hybrid subject, part of mathematics and different from subjects in physics, but distinct from arithmetic and geometry in possessing a physical character.

Ampère (1834) presented a rather detailed and elaborate classification scheme for the sciences in his *Essai sur la philosophie des sciences* of 1834. The mathematical sciences were made up of arithmetic, geometry, mechanics, and astronomy. Arithmetic and geometry were mathematical subjects "proprement dits," while mechanics and astronomy were physico-mathematical in character. The physical sciences included atomic theory and chemistry.

Ampère's point of view was reflected in some later French writers on scientific classification. Thus Renouvier (1859) in his *Essais de critique générale* put rational mechanics and applied mathematics together with mathematical subjects (arithmetic, algebra, mathematical analysis, geometry) in the category of Logical Sciences, which were to be distinguished from Physical Sciences, the latter including astronomy. At the end of the century Goblot's *Essai sur la classification des sciences* (1898) positioned mechanics as part of mathematics and distinguished it from physics. The estimable Scottish authority Flint (1904, 278), in his survey of work on classification, seemed to regard mechanics as part of mathematics, writing that mechanics "is as abstract as Geometry, and in its applications is not more concrete," and "Mechanics is both abstract and concrete, both quantitative and qualitative, and cannot be denied to be on the borderland between mathematical and physical science."[3]

Although Ampère's understanding of the scope of mathematics was adopted by some later authors, the view that came to be much more common as the century progressed was just the opposite. There was a decided shift away from the view that mechanics was part of mathematics. In the last part of the century both humanistic and scientific thinkers interpreted mathematics as a subject more or less co-extensive with what today would be called pure mathematics.

[3]Flint (1904, 222–223 and 308–312) gives accounts of the classifications of Renouvier and Goblot.

This shift is apparent in the writings of the English polymath Herbert Spencer (1864), who published in 1864 *The Classification of the Sciences*. Spencer opposed Comte's hierarchy, mainly on the grounds of the reductionist ordering of the sciences along a linear sequence that it implied. Spencer was among that group of thinkers who believed that logic and mathematics were closely connected and distinguished by their abstractness from the natural sciences. Mathematics and logic dealt with relations, while the natural sciences dealt with objects. Rather than putting the natural sciences into a sequence he divided them into two distinct groups: the abstract-concrete sciences, consisting of mechanics, physics and chemistry; and the concrete sciences, consisting of astronomy, geology, biology, psychology, and sociology.

The exclusion of mechanics from mathematics was also advocated by the Austrian physicist Ernst Mach, who published his noted critical and historical account of mechanics in 1883. Although Mach's philosophy shared similarities with the empiricist outlook of Comte and Ampère, he insisted that mechanics was not part of mathematics. At the beginning of the preface to his book he proclaimed:

> Mechanics will here be treated, not as a branch of mathematics, but as one of the physical sciences. If the reader's interest is in that side of the subject, if he is curious to know how the principles of mechanics have been ascertained, from what sources they take their origin, and how far they can be regarded as permanent acquisitions, he will find, I hope, in these pages some enlightenment. All this, the positive and physical essence of mechanics, which makes its chief and highest interest for a student of nature, is in existing treatises completely buried and concealed beneath a mass of technical considerations. (Mach 1883, i)[4]

Mach was influenced by his phenomenological understanding of mechanics and his belief that *a priori* metaphysical conceptions had no place in physics, a mistake that could arise if mechanics was taken as part of mathematics. There were also important developments in nineteenth-century mathematics that influenced scientific thought in the century's second half. In a footnote toward the end of his book Mach discussed the discovery of non-Euclidean geometry. This discovery showed that geometry was not simply a description of spatial reality, for there were multiple geometries and only one spatial reality. Mathematics including geometry was evidently about intellectual structures, while mechanics was about objects in the external world. Non-Euclidean geometries existed, but non-inertial physics did not. Mach was opposed to the interpretation of the properties of real space ("die Eigenschaften des gegebenen Raumes") by what he called "the pseudo-theories of geometry that seek to excogitate these properties by metaphysical arguments."

The common view among the classifiers of science in the second half of the century was that mathematics did not include mechanics. This fact is apparent in a broad range of authors discussed by Flint in his 1904 historical survey. Whewell in 1858 distinguished mathematics (arithmetic, geometry, algebra, differentials) from

---

[4]English translation is by Thomas J. McCormack from the 1893 English edition of Mach's book, *The Science of Mechanics; A Critical and Historical Account of Its Development* (Open Court, Chicago).

astronomy and mechanics (Flint 1904, 198). Wilson in 1856 separated mechanics, which he called a practical science, from mathematical subjects (arithmetic, algebra, geometry, calculus) which made up, with the study of method and ontology, the pure sciences (Flint 1904, 215–216). de Roberty in an 1881 book on sociology separated mathematics from mechanics, regarding the latter as a descriptive science (Flint 1904, 263–4). In his 1887 book *Versuch einer concreten Logik*, the Prague philosopher Tomáš Masaryk (1887) advocated a hierarchal conception of science, placing mathematics first and assigning mechanics to a second group (Flint 1904, 277–8).[5] Masaryk followed Mach in explicitly separating mechanics from mathematics. In 1870 the Scottish philosopher Alexander Bain asserted that mathematics was distinct from mechanics, and placed the latter with physics (Flint 1904, 241–2). In Pearson's *Grammar of Science* of 1892 logic and mathematics were classified as abstract sciences, while mechanics was one of the concrete sciences. One year later Raoul de la Grasserie (1893) followed Herbert Spencer in classifying mathematics as an abstract science and mechanics as abstract-concrete (Flint 1904, 289–292). Writing in the early 1930s but expressing long-held views, Bliss (1933, 293) asserted that the possibility of a mathematical treatment of mechanics "should not mislead scientists to admit the claims of some mathematicians that Mechanics is merely a branch of Mathematics. That is not true even of Rational, or Analytic Mechanics, which of course should not be dissevered from the sub-science as a whole."

The book-classification schemes at the end of the century were united in limiting the scope of mathematics, and in either placing mechanics within physics or including it as a subject area in its own right. In Cutter, mechanics was put with physics rather than mathematics, while astronomy was made a distinct subject area, after chemistry. Although Dewey (1886) had included some applied subjects in mathematics, mechanics was placed in physics, along with optics, thermodynamics, and electromagnetism. In the International Catalogue of Scientific Literature (see §4) mechanics received its own subject area, intermediate between mathematics and physics. In the systems of both Brown and Bliss, mechanics is separated from mathematics and classified as a physics subject, along with thermodynamics and electromagnetism.

Alone among the major classification systems, the LC scheme placed mechanics under mathematics, and situated astronomy as a subject field between mathematics and physics. The title of the original LC volume on mathematics is worded "Class QA: Mathematics (Including Analytic Mechanics)." It is not entirely clear why the architects of LC proceeded this way, but the grouping of mechanics within mathematics is a singular feature of the LC classification system that continues to the present.

[5]Flint (1904, 277) mistakenly gives the date of publication of Masaryk's book as 1866. Masaryk was born in 1850 and entered the University of Vienna in 1872.

4 The Place of Calculus/Analysis in Classification Schemes for Mathematics

Comte's distinction between abstract mathematics consisting of arithmetic, algebra and calculus, on the one hand, and concrete mathematics, consisting of geometry and mechanics on the other, reflected a classificatory order that placed calculus ahead of geometry. It was also in keeping with the prevailing conception in French mathematics of calculus as a form of "algebraic analysis," the very title of Augustin Cauchy's famous textbook of 1821 on the calculus.

In his 1834 book Ampère introduced neologisms to designate the various subject areas of mathematics. What he called "arithmologie" was divided into two parts, the first consisting of arithmetic and algebra, and the second consisting of the theory of functions and the theory of probabilities. The theory of functions encompassed calculus-related parts of mathematics. Geometry was the second subject area of mathematics, under which Ampère placed synthetic and analytic geometry, as well as the theory of lines and surfaces and something called molecular geometry. The last subject area of mathematics consisted of the physico-mathematical subjects mechanics and astronomy (the latter called "Urinologie" by Ampère.) Mechanics in turn was divided into elementary and transcendental mechanics, while astronomy was divided into general astronomy and celestial mechanics.

Among the many writers who wrote on classification of science from the 1840s to the end of the century, the predominant tendency was to depart from Comte and Ampère by placing geometry ahead of calculus. Mathematical subjects were placed in the standard order: arithmetic, algebra, geometry, and calculus. Whewell in 1840 conceived of mathematics as the subjects "Geometry, Arithmetic, Algebra, and Differentials, and based on the ideas of space, time, number, sign, and limit" (Flint 1904, 199). Bain in 1870 divided mathematics into arithmetic, algebra, geometry, algebraic geometry, and the higher calculus (the latter dealing with incommensurable magnitudes) (Flint 1904, 199). Wilson in 1856 gave the order arithmetic, geometry, algebra, calculus, trigonometry, and analytic geometry (Flint 1904, 216). Janet in 1897 used abstraction as something that distinguished arithmetic, geometry and mechanics from algebra and the differential and integral calculus (Flint 1904, 304). Flint (1904, p. 278) himself wrote that "Arithmetic and Geometry are very different both as to matter and method from Calculus and Kinematics."[6]

[6]An exception to the prevailing consensus was Karl Pearson, who in his *Grammar of Science* (1892) put theory of functions and calculus together with arithmetic and algebra, these subjects dealing with quantity, while geometry was classified as a distinct subject area dealing with space (Flint p. 296). Earlier the Paris book seller Jacques-Charles Brunet (1814, 1860) in his pioneering classification scheme placed mathematical subjects in the order arithmetic, algebra, calculus, and geometry. Brunet was presumably influenced by Comte and Ampère. Brunet's catalogue was exceptional among all classification schemes in placing mathematics at the end of the sciences, following philosophy, physics, chemistry, geology, biology, and medicine.

With the exception of the Library of Congress, the major library classification schemes around 1900 placed geometry before calculus. Dewey and Cutter both adopted the order arithmetic, algebra, geometry, trigonometry, and calculus, while Brown presented these subjects in the order arithmetic, algebra, geometry, calculus, and trigonometry. The librarians presumably were guided by historical and pedagogical considerations: calculus had originated as a set of methods for the study of curves and surfaces, and calculus was a more advanced teaching subject than elementary geometry and therefore was placed after it. The librarians may also have perceived the natural order to be one of successive abstraction, and calculus and higher analysis were viewed as more abstract than geometry.

Although the subject of the present essay is the classification of books, it is necessary to look at how periodical mathematical literature was classified by subject in the second half of the nineteenth century, as this would bear directly on the classification scheme for mathematical books adopted by the LC. Unlike book classification, which was aimed at a very broad readership at various levels of engagement with the subject, the practices followed by journals reflected the outlook of advanced researchers in the field. Insofar as the ordering of subjects is concerned, the point of view was essentially a continuation of the French outlook expressed by Comte and Ampère early in the century. The *Zeitschrift für Mathematik und Physik*, founded in 1856, was one of the first journals to explicitly divide its contents into subject categories. The latter were presented in the order arithmetic and analysis, geometry, mechanics, optics, electricity and Galvinism, and smaller and miscellaneous subjects. The grouping of analysis with arithmetic and its placement ahead of geometry reflected the prevailing view of advanced researchers, and indicated more generally the well-known "arithmetization of analysis" in the development of mathematics in the nineteenth century. Calculus in its original formulation was known as "fine geometry," and such eighteenth-century masters of analysis as Euler and Lagrange were known as geometers. By the second half of the nineteenth century the research picture had shifted radically, and geometry had become something of a subsidiary subject with respect to the primary grounding of advanced mathematics in arithmetic and analysis.

Carl Ohrtmann and Felix Müller were Berlin gymnasium teachers of mathematics who founded in 1871 the abstracting periodical *Jahrbuch über die Fortschritte der Mathematik*. There was a large increase in the growth of mathematical literature in the nineteenth century, and a corresponding need to assist researchers in navigating materials published in their fields. Ohrtmann and Müller modelled the *Jahrbuch* after an abstracting journal for physics that had already been in existence for close to twenty-five years, the *Fortschritte der Physik/Physikalische Berichte*. Although the publications reviewed in the *Jahrbuch* consisted mainly of periodical literature, books were also included. Subjects were presented in the order history and philosophy, algebra, number theory, series, differential and integral calculus, function theory (complex functions), pure, elementary and synthetic geometry, analytic geometry, mechanics, mathematical physics (electromagnetism, theory of heat, optics), and geodesy and astronomy.[7] Since there was already a physics abstracting journal, the physics subjects included in the *Jahrbuch* were ones in which the treatment was very mathematical.

At the end of the century the Royal Society of London established the International Catalogue of Scientific Literature A Mathematics (1902), a major international bibliographic project that was intended to cover both periodical and book literature. In this work mathematics (which was also referred to as "pure mathematics") was divided into the following subject areas: fundamental concepts, algebra and number theory, analysis, and geometry. This ordering of subjects became canonical in the classification of twentieth-century mathematical literature, at least as this was followed by the LC and mathematical abstracting services. (It should be noted that the Dewey system continued to place geometry before calculus and analysis up until the late 1960s, at which time its schedules were revised and brought into alignment with the LC.)

The classification schedules for mathematical subjects in the original LC system of 1905 were compiled by J. David Thompson, chief of the science section, under the direction of Martel, head of classification for the whole of LC (See Library of Congress Classification Class Q Science (1905).). Thompson was a native of England who had studied mathematics at the University of Cambridge, graduating 16th Wrangler in 1895. In the preface to the volume on science he (1905, 3) states that he has relied notably on the schedules of the International Catalogue of Scientific Literature. While the overall scheme of the LC system was patterned on the Cutter system of classification, the organization of scientific subjects followed the ICSL. Insofar as advanced mathematical subject areas were concerned, Thompson followed the ICSL very closely. The 1905 edition of the LC science schedules was republished in multiple later editions, each containing modifications and extensions of the original scheme.

In the 1930s there were two new library classification systems, Bliss's bibliographic classification and Ranganathan's (1933) colon classification. Although Bliss presented the three subject areas of mathematics as arithmetic-and-algebra, geometry and analysis, for the purposes of classification he placed them in the order arithmetic-and-algebra, analysis, and geometry. He made this change for reasons of what he called "collocation," apparently referring to the usage established by the ICSL and the LC. Ranganathan also classified mathematics subjects in the order arithmetic, algebra, analysis, and geometry, and followed LC in including mechanics within mathematics. In a departure from all other classification schemes he placed astronomy within mathematics.

[7] Göbel (2008, 9–13) observes that "Später bildeten diese Abschnitte die Grundlage für den Aufbau einer Mathematik-Klassifikation." ("These divisions later formed the basis for the construction of a mathematics classification.")

5 Analysis in the LC Classification for Mathematics: The Case of Functions of a Complex Variable

In the LC classification system books on science are classified under Q, and those on mathematics are classified under QA. In 1905 some parts of mathematics hardly existed yet as recognized subject areas. In the ICSL under arithmetic there was a subject entry on "aggregates," what would later be called the theory of sets, but there was no entry at all for this subject in the LC. When Abraham Fraenkel's *Einleitung in die Mengenlehre* appeared in 1919 it was classified in the LC under foundations of arithmetic (QA248) in the algebra section, and that became the standard LC subject classification for books on set theory.

A part of mathematics that was very well established in 1905 was analysis, and books on this subject received call numbers in the range from QA300 to QA400. The theory of functions was designated QA331 and was made up of books we would regard today as belonging to complex analysis. The theory of functions of a real variable came to be designated QA331.5, being regarded as a branch or offshoot of the theory of functions. The classification scheme is evident in the following two books on analysis from the early years of the century:

QA331 Heinrich Burkhard, *Theory of Functions of a Complex Variable* (1913)

QA331.5 James Pierpont, *Lectures on the Theory of Functions of Real Variables* (1905–12)

When Lars Ahlfhors' *Complex Analysis* was published in 1953 it was given the LC subject designation QA331. In the 1960s complex analysis replaced the theory of functions as the standard subject name for the theory of functions of a complex variable (Figure 1).[8] At this time one also began to see the publication of books

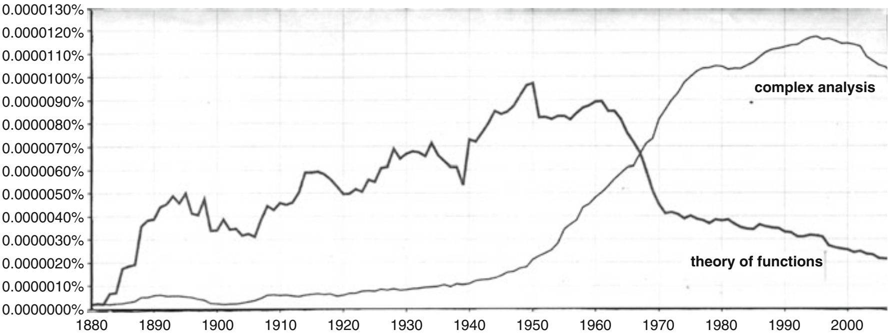


Fig. 1 Google Ngram: theory of functions, complex analysis

[8]An interesting graphical illustration of this change in usage is provided by the Google Ngram (Michel et al. 2011) for the frequency of the terms "theory of functions" and "complex analysis" for the period from 1880 to 2008. See Figure 1.

with the term "real analysis" in the title. H.L. Royden's *Real Analysis* appeared in 1963 and was given the subject designation QA331.5. Thus real analysis was envisaged in this classification scheme as an offshoot of complex analysis. The earlier subject classifications QA331 (theory of functions, implicitly functions of a complex variable) and QA331.5 (theory of functions of a real variable) mapped onto the new subject names complex analysis (QA331) and real analysis (QA331.5).

In the LC books on analysis with the classification QA300 are devoted to the more general parts of analysis and the foundations of the subject. A widely used primer on analysis for senior undergraduate and graduate students from the 1950s and 1960s was Walter Rudin's *Principles of Analysis* (1953 and later editions). Rudin's book was classified under QA300. We have the classification sequence:

Q Science
  QA Mathematics
    QA300 Rudin *Principles of Analysis*
      QA331 Ahlfors *Complex Analysis*
        QA331.5 Royden *Real Analysis*

By the 1970s some books on real analysis were assigned the designation QA300, and thus were understood to belong to more general parts of analysis, prior in the classification scheme to complex analysis. Other books on real analysis continued to receive the traditional designation QA331.5. There was an overhaul of LC mathematical analysis subject designations in the 1980s, and this change is contained in the current schedule that may be found online (http://www.loc.gov/aba/publications/FreeLCC/Q-text.pdf). The change had been made by around 1990. Here is how the breakdown for subjects in analysis is now given:

Analysis
QA300 General works, treatises, and textbooks

Theory of functions
QA 331 General works, treatises, and advanced textbooks
QA 331.3 Elementary textbooks
QA 331.5 Functions of real variables
QA 331.7 Functions of complex variables
  Riemann surfaces including multiform, uniform functions

Evidently the QA331 section dealing with the theory of functions has been reorganized to reflect the standard order of subject presentation: first general works, followed by elementary presentations, and then according to some presumably natural principle, an ordered list of the subject areas that fall under the theory of functions.

An old principle of book classification followed by the LC that is very useful to the historian is that books are not reclassified when a revision, either major or minor, of the classification system takes place.[9] This seems to be partly for practical reasons—it would be difficult for libraries to be continually reclassifying the materials in their collections. It should be noted that although the classification of a book is not changed, in the LC a later edition of a given book may have a different call number. For example, Stanley G. Krantz's *Function Theory of Several Complex Variables* was classified as QA331 when it appeared in 1982, a designation that remains unchanged to this day, while the second edition of this book in 1992 received the call number QA331.7.

The principle of the organization of scientific subjects which is followed in library catalogues is foundational: more basic subjects come first, followed by progressively more complex subjects. Underlying the conception of foundation is a building metaphor, invoking the structure and construction of a building. A classification where real analysis is placed before complex analysis is consistent with a foundational conception of subject classification. In the original LC classification, where the theory of functions of a real variable is a sub-subject of the theory of functions, David Thompson was presumably thinking of classification in a somewhat different way, as a division in which the complete subject comes first, and where one proceeds from there to obtain various special subject areas that fall within the general subject. In certain contexts this second approach to classification may seem more natural or practical, as it would, for example, if one were organizing goods in a department store. To find a given make of coffee maker one would locate the section on household goods, proceed to the section on kitchen supplies, and then find the section on kitchen appliances, ending finally in the section on coffee makers. The conventional ordering of intellectual subjects—at least ones in science—follows a different, foundational principle that is inherent in the epistemological character of the subject matter.

It should be noted that there was also an evolution in the classifications schedules for mathematics employed in the Dewey Decimal system. The original Dewey schedule from 1885 for books on mathematics was:

511 Arithmetic
512 Algebra
513–516 Geometry
517 Calculus (analysis)

[9]While there is a general conservatism among librarians with respect to classification, in the case of the Dewey Decimal system there have been revisions of the classification that have been retroactively applied by some libraries to books in their collections. For example, before 1970 books on analysis were classified by Dewey under 517 (after geometry), whereas after 1970 such books were assigned the classification 515 (before geometry). In the public libraries of Cleveland and Cincinnati the older books retain their original classifications. However, in the library of the University of Illinois at Champaign-Urbana, where the Dewey system is used, the older books on analysis have been assigned the new classification numbers. This is also generally true of the Toronto Public Library.

Around 1970 the Dewey classification for books on analysis was changed from 517 to 515, while books in geometry that originally would have been classified under 515 were classified as 516. For example, a book on complex analysis published in 1965 was classified under 517, while a book on the same subject by the 1970s was classified under 515. In addition, the classifications for arithmetic and algebra were reversed, so that arithmetic became 513 while algebra remained at 512. The current Dewey classification schedule for books in mathematics is:

512 Algebra
513 Arithmetic
514 Topology
515 Analysis
516 Geometry

The change in the classification brought the Dewey system into conformity with the LC as well as with the classifications used by abstracting periodicals such as *Mathematical Reviews*.

Prior to about 1970 in the Dewey system books on functions of a complex variable and books on functions of a real variable where both standardly catalogued under the call number 517.5. There was no relative placement of one group with respect to the other, as there was in LC. When books with complex analysis and real analysis in their titles appeared in the 1950s and early 1960s they were all given the same call number 517.5. During the later 1960s books on complex analysis were being classified as 517.8, while ones on real analysis continued to receive 517.5. After 1969 or 1970 all books on analysis shifted from 517 to 515, with books on real analysis being assigned 515.8, and ones on complex analysis being assigned 515.9.[10]

Acknowledgements I would like to thank Hardy Grant for advice on an earlier draft of this paper and Matthew Edwards for helpful discussions. I am grateful to two referees for their comments on the paper. I am also grateful to Noah Stemeroff for help with LaTeX and for his careful proof reading of the paper.

References

Ampère, André-Marie. (1834) *Essai sur la philosophie des sciences, ou exposition analytique d'une classification naturelle de toutes les connaissances humaines. Tome 1*. Mallet-Bachelier, Paris.

Bain, Alexander. (1870) *Logic Part First Deduction*. Longmans, Green, Reader and Dyer, London.

Bliss, Henry E. (1929) *The Organization of Knowledge and the System of the Sciences*. Henry Holt and Company, New York.

[10]There was not complete consistency in cataloguing during the transition years between about 1968 and 1972. It should also be noted that books on real analysis were sometimes assigned 517 rather than 517.5 (before 1970) and 515 rather than 515.8 (after 1970).

Bliss, Henry E. (1933) *The Organization of Knowledge in Libraries and the Subject-Approach to Books*, H.W. Wilson Company, New York.

Bliss, Henry E. (1935) *A System of Bibliographic Classification*. The H.W. Wilson Company, New York.

Brown, James Duff. (1906) *Subject Classification, with tables, indexes, etc. for the sub-division of subjects*. Library Supply, London.

Brunet, Jacques-Charles. (1814, 1860) *Manuel du libraire et de l'amateur de livres*. Dido Frères, Paris. The first edition of this work was published in 1814. An English translation and introduction by Donald Bruce McKeon of the 1864 edition appeared in 1976 as No. 1 of the Louisiana State University Graduate School of Library Science Occasional Papers. Baton Rouge. In an appendix the translator gives a statistical breakdown of the contents of Brunet's bookshop, using the total number of bench marks or call numbers each subject receives as a percentage of the total number of bench marks overall. Arts and Sciences constituted 22.7% of the contents of the collection, while the mathematical sciences constituted 15.2% of Arts and Sciences. From this one may deduce that the mathematical sciences represented 3.5% of the overall collection.

Brunet, Jacques-Charles. (1860, 1864) *Manuel du libraire de l'amateur de livres. Tome premier*. Librairie de Firmin Didot Frères, fils et Cie. Five more volumes appeared up to 1865.

Comte, Auguste. (1830) *Cours de philosophie positive tome premier Contenant les préliminaires généraux et la philosophie mathématique*. Rouen Frères, Paris.

Cutter, Charles A. Boston (1882) *Athenaeum; How to Get Books, With an Explanation of the New Way of Marking Books*. Rockwell Churchill, Boston.

Cutter, Charles A. (1891–93) *Expansive Classification Part I: The First Six Classifications*. Boston.

de la Grasserie, Raoul. (1893) *De la classification objective et subjective des arts, de la littérature des sciences*. F. Alcan, Paris.

de Roberty, Eugène. (1881) *La Sociologie*. Germer Baillière, Paris.

Dewey, Melvil. (1886) *A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library*. Case, Lockwood & Brainard Company, Amherst, Mass.

Dolby, R.G.A. (1979) "Classification of the Sciences The Nineteenth Century Tradition," in (Roy Ellen and David Reason, 1979), pp. 167–193.

Ellen, Roy F. and David Reason. (1979) *Classifications in their Social Context*. Academic Press, London, New York and San Francisco.

Erdlund, Paul.(1976) "A Monster and a Miracle: The Cataloging Distribution Service of the Library of Congress, 1901–1976," *The Quarterly Journal of the Library of Congress* 33, pp. 383–421.

Flint, Robert. (1904) *Philosophy as Scientia Scientiarum and History of Classifications of the Sciences*. William Blackwood and Sons, Edinburgh and London.

Göbel, Silke. (2008) "Die Anfänge des Wissensmanagements in der Mathematik: das „Jahrbuch über die Fortschritte der Mathematik"", in Bernd Wegner (Ed.), *Mathematik in Blickpunkt*, pp. 9–13. FIZ Karlsruhe.

Goblot, Edmond. (1898) *Essai sur la classification des sciences*. Germer Baillière, Paris.

Hacking, Ian.(2014) *Why Is There Philosophy of Mathematics At All?* Cambridge University Press.

International Catalogue of Scientific Literature A Mathematics. (1902) Royal Society of London, London.

Janet, Paul. (1897) *Principes de métaphysique et de psychologie Leçons professées à la Faculté des Lettres de Paris 1888–1894 tome 1*. Librairie Ch. Delagrave, Paris.

LaMontagne, Leo E. (1961) *American Library Classification, with special reference to the Library of Congress*. Shoe String Press, Hamden, Conn.

Library of Congress Classification Class Q Science. (1905) Government Printing Office, Washington.

Mach, Ernst. (1883) *Die Mechanik in ihrer Entwickelung historisch-kritisch dargestellt*. Brockhaus, Leipzig.

Masaryk, Tomáš. (1887) *Versuch einer concreten Logik*. Verlag von Carl Konegen, Vienna.

Michel, Jean-Baptiste et al. (2011) "Quantitative Analysis of Culture Using Millions of Digitized Books," Science Vol. 331 (Issue 6014), pp. 176–182.

Pearson, Karl. (1892) *Grammar of Science*. Walter Scott, London.
Ranganathan, S.R. (1933) *Colon Classification*, Madras Library Association, Madras, E. Goldston, London.
Renouvier, Charles. (1859) *Essais de critique générale*. A. Colin, Paris.
Sayers, W.C. Berwick. (1915–1916) *Canons of Classification*. Grafton and Co., London and The H.W. Wilson Company, White Plains, N.Y.
Siegel, Max H. (2016) Book review of (Hacking, 2014). *Mind* 125 (497), pp. 252–6.
Spencer, Herbert . (1864) *The Classification of the Sciences*. D Appleton and Company, New York.
Whewell, William. (1858) *Novum Organon Renovatum: Being the Second Part of the Philosophy of the inductive Sciences*. J.W. Parker, London.
Wilson, William D. (1856) *Elementary Treatise of Logic*. D. Appleton and Company, New York.

The Convolution as a Mathematical Object

Roger Godard

Abstract For many scientists, the convolution is well known as a tool of applied mathematics dating from the 20th century. However, the idea of a discrete convolution can be traced as far back as the Middle Ages in China. After d'Alembert and Euler, applications of convolution integrals appeared at the end of the 18th century and the beginning of the 19th century. These were in potential theory, the heat conduction equation, and the wave equation as developed by Lagrange, Laplace, Legendre, Fourier, Cauchy, and Poisson. Here we emphasize Fourier's contribution to the superposition principle for memory, time delay, and superposition of events. We consider the modern mathematical properties of the convolution with Volterra, Lebesgue, Doetsch, and Schwartz, and we illustrate some applications of the convolution in the theory of probability, interpolation, and smoothing.

Keywords Convolution integrals • Gravitational theory • Fourier series • Superposition principle • Theory of distributions • Probability theory • Filter theory

1 Introduction

Integral equations have been well studied and documented by historians of mathematics; for example, see Lonseth (1977), Kline's *Mathematical Thought* (1972 Vol. 3: 1052-1095), Dieudonné's *History of Functional Analysis* (1981: 97-105), or Bourbaki's *Éléments d'histoire des mathématiques* (1960: 233-239). However, until recently, the full history of the convolution integral has been ignored.

Alejandro Dominguez-Torres (2010, 2015) has surveyed the origin and history of the convolution integral, and Gardner and Barnes (1942: 364-365) have also presented valuable comments on the history of the Laplace transform and the convolution integral. We have chosen instead to focus this present work towards

R. Godard (✉)
Department of Mathematics and Computer Science, Royal Military College of Canada,
92 Florence Street, Kingston, ON, Canada K7M 1Y6,
e-mail: rgodard3@cogeco.ca

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*,
Proceedings of the Canadian Society for History and Philosophy of Mathematics
La Société Canadienne d'Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_13

mathematical physics. Because the convolution has been a basic mathematical tool for more than two hundred and thirty years, it is impossible to survey all of its possibilities. The convolution integral (CI) is defined as follows:

$$(f^{*}g)(x) = \int_{-\infty}^{+\infty} f(x-t)\, g(t) dt, \tag{1}$$

where the result is a function of the variable x only. The symbol *, already in use during the 1920s, was chosen to emphasize that a convolution product is not an ordinary product. The transformation of $f(t)$ into $f(-t)$ is called folding in English from the word *Faltung* in German; $(x - t)$ is a translation kernel. The old French expression for this integral was *produit de composition*. We shall present three important properties of the convolution: (a) its use in expressing the concept of memory or time delay in mathematics, (b) the smoothing properties of $f{*}g$ with respect to the functions $f(t)$ and $g(t)$, and (c) its additional mathematical properties.

2 Convolution and Mathematical Physics

Under the modern definition, a convolution corresponds to the mathematization of a memory problem, smoothing processes or, more precisely, it involves an operator with translational invariance, and the famous principle of causality that the future cannot influence the present (Sirovich 1988: 80-82). During the 17th century, physical properties of the convolution appeared implicitly in Huygens' *Treatise of Light* (Huygens 1690: 17, Godard 2010), written during his stay in Paris and presented to the Academy of Sciences in 1678. But the major breakthrough in mathematical physics and convolution integrals would have to wait until the birth of the Newtonian gravitational theory. In Newton's Principia of 1687, the gravitational forces considered are the forces of attraction acting upon two material points. If we denote by μ_1 a mass situated at the point M_1 and by μ_2 the mass situated at the point M_2, the mutual force of attraction is expressed by:

$$\vec{F} = \gamma \frac{\mu_1 \mu_2}{r^2} \vec{e_r}, \tag{2}$$

where r is the radial distance between M_1 and M_2, γ is a constant, and $\vec{e_r}$ is a unit vector indicating the direction from one mass to the other.

After Maupertuis' observations about the sphericity of the earth in 1736, the theory of the shape of the earth became a topic of interest. Clairaut published his famous book about the form of the earth in 1743. For him, such a study of the earth was a step towards and understanding of the system of the world. An object, put in a given place outside the earth was attracted by all particles composing the earth, each of them acting according to its position and simultaneously with more or less force according to its distance. Clairaut went so far as to consider the earth as composed of inhomogeneous matter. Geometers soon realized that it was difficult to manipulate a vector force, and began to work with its components individually.

In 1748, Daniel Bernoulli first introduced in fluid mechanics the concept of a scalar *potential function* from which a force is derived, and in 1773, Lagrange presented the scalar *quantity* Ω for the problem of a body attracted by a system of point masses (Lagrange 1773):

$$\Omega = \frac{m_1\mu}{r_1} + \frac{m_2\mu}{r_2} + ..\cdots = \frac{M}{\Delta} + \frac{M^{'}}{\Delta^{'}} + \frac{M^{''}}{\Delta^{''}} + \ldots, \tag{3}$$

where r_i or Δ^i (Lagrange's notation) are the distances from points of mass m_i of coordinates (x_i, y_i, z_i) to the point of mass μ of coordinates (a, b, c), and the M^i (Lagrange's notation) is $m_i\mu$. In a more modern notation, we can express the above as:

$$V(a, b, c) = \sum_i \frac{m_i\mu}{\sqrt{(a-x_i)^2 + (b-y_i)^2 + (c-z_i)^2}}. \tag{4}$$

We shall see that this latter expression is a discrete convolution formula. Note that the term *potential function V* itself was not proposed until much later by Green in his 1828 paper "An Essay on the Application of Mathematical Analysis to the theories of Electricity and Magnetism" and in 1840 Gauss wrote it as a *potential* only. The next stage was to generalize the formula from the discrete case to the continuous case with Laplace, Legendre, and Lagrange. In 1792-1793, Lagrange wrote the following statements:

> "We know that the attraction of the spheroid on any point where the position in space would be determined by the coordinates a, b, c brought to the same axis as the coordinates x, y, z depends upon the formula:
>
> $$\int \frac{dxdydz}{\sqrt{(a-x)^2 + (b-y)^2 + (c-z)^2}}, \tag{5}$$
>
> which I call V, the integration being brought to the entire mass of the spheroid. Such that, if in the quantity V seen as a function of $a, b, c \ldots$" (Lagrange 1798).

Let us compare (1) to (5), and choose in (1):

$$g(t) = 1, \text{and } f(a-x, b-y, c-z) = \frac{1}{\sqrt{(a-x)^2 + (b-y)^2 + (c-z)^2}} \tag{6}$$

In this present work, we shall encounter several convolution integrals where the function $g(t)$ of (1) is unity. Equation (5) is the first three-dimensional convolution integral of mathematical physics. In (5), the earth is assumed to be a homogeneous body and the mass is taken as unity. Then, Laplace showed that (5) was a solution of the Laplace equation $\nabla^2 V = 0$. The formula occurred several times in Laplace's

Traité de mécanique celeste (1798-1823). In 1813, Poisson observed that (5) diverged at internal points of interest and that the Laplace equation had to be modified. Through the work of Green, Gauss, and Poincaré, the potential theory remained a strong field of interest throughout the 19th century.

3 Convolution and Partial Differential Equations of Mathematical Physics

We showed in the last section that a convolution integral (CI) could be the solution of partial differential equations, and its importance as mathematical object would be affirmed at the beginning of the 19th century. Laplace[1] (1809) published a work on a diffusion equation with the results expressed as a convolution-like integral. Let us quote the assessment of Grattan-Guiness (1972):

"*Fourier's reliance on Laplace's results was not so complete, but the analysis gave him the clue which he had been looking: instead of a series solution, he had found a* (convolution) *integral solution.*" Convolution integrals first appeared in Fourier's 1811 paper (Deakin 1981: 362), and within a few years, these integrals blossomed in the works of Fourier, Cauchy, and Poisson. Fourier (1822) truly mastered this new mathematical object and we have found more than forty convolution integrals in the *Théorie analytique de la chaleur* (1822).

In his note concerning vibrations of elastic surfaces, Fourier (1818) studied the equation of the vibrating membrane:

$$\frac{\partial^2 v}{\partial t^2} + \frac{\partial^4 v}{\partial x^4} + 2\frac{\partial^4 v}{\partial x^2 \partial y^2} + \frac{\partial^4 v}{\partial y^4} = 0, \tag{7}$$

and its solution as a convolution:

$$v\,(x, y, t) = \frac{1}{l}\int d\alpha \int \varphi\,(\alpha, \beta)\sin\frac{(x-\alpha)^2 + (y-\beta)^2}{4t}d\beta, \tag{8}$$

where $\varphi(\alpha, \beta)$ represents the initial state, v is the position of the membrane, and l is the length of the membrane. In the same article Fourier makes the following comparison with the wave equation:

[1]Note that several convolution integrals previously appeared in the 1805 Lacroix "Traité des différences et des séries" as the solutions of partial differential equations. For example, we found on page 506 the following convolution integrals: $\xi(u, v) = \int T(u-t)\varphi(t)dt + \int T_i(v-t)\psi(t)dt$ and a double integral on page 518: $\iint \frac{(u-x)^n(v-y)^m}{1.2.3...n.1.2.3...m}\psi\,(x, y)\,dxdy$. Lacroix's book is available from the website of the Bibliothèque Nationale de France.

$$\frac{\partial^2 v}{\partial t^2} + \frac{\partial^2 v}{\partial x^2} = 0, \tag{9}$$

where the integral solution is:

$$v\,(x,t) = \frac{1}{\sqrt{\pi t}} \int_{-\infty}^{+\infty} \varphi\,(\alpha) \sin\left[\frac{\pi}{4} + \frac{(x-\alpha)^2}{4t}\right] d\alpha. \tag{10}$$

and φ(α) is the initial state. This would prove to be his most important article for the understanding of a convolution integral. Fourier's objectives were ultimately to prove that these integrals could describe complex natural effects. Fourier presents solutions of the heat diffusion equation, the wave equation at the surface of a liquid, and proceeds to the question of elastic blades under the structure of convolution integrals. Struck by the similarity between the solutions, that in each equation the initial state and time were included; Fourier would remark:

"*The analogy from which we speak does not result of the physical nature of causes; it resides all together in the mathematical analysis that lends common forms to most diverse phenomena*" (Fourier 1818).

The secret was a hidden property of convolution integrals. The present state of the temperature of an infinite prism, of the position of an elastic blade, or of the state of a wave depends upon their past history. In the Newtonian potential theory, geometers obtained a spatial convolution, while in time-dependent problems they obtained time-dependent convolution integrals. The same phenomenon can be found in Abel's 1823-1826 solution of his problem of mechanics (Kline Vol. 3: 1052-1054): a particle slides from point P to point *Q* down a smooth curve, and the transit time depends upon its past trajectory.

As previously stated, Fourier used convolution integrals in his 1822 *Théorie analytique de la chaleur*. Considering the case of a three-dimensional solid where heat is propagated in all directions. Fourier (1822: 393) found the solution of the variable temperature *v* as a triple integral:

$$v\,(x,y,z,t) \quad = \quad \int_{-a_1}^{a_2} da \int_{-b_1}^{b_2} db \int_{-c_1}^{c_2} dc \frac{e^{-\frac{(x-a)^2+(y-b)^2+(z-c)^2}{4t}}}{2^3 \pi^{\frac{3}{2}} t^{\frac{3}{2}}} f\,(a,b,c)\,, \tag{11}$$

where the function $f(a,b,c)$ represents the temperature of the solid which was initially heated, and $a_1\,, a_{2,} b_1\,, b_2\,, c_1\,, c_2$ represent the dimensions of the solid. Fourier was in no way intimidated by these integrals. Note that already, in E274 published in (1763), Euler had proposed an integral solution to an ordinary differential equation $y(u) = \int P(x)(u+x)^n dx$ where the function $P(x)$ comes from the differential equation itself. Regarding this expression, Euler wrote (Deakin 1985):

"*When this function P becomes known, the integration is performed, at least by quadrature, for each value of u, which during the integration is like a constant.*"

With the development of Fourier trigonometric series, greater studies of their nature became fundamental (Fourier 1822: art. 415-417, Poisson 1823, Lejeune-Dirichlet 1829, Sachse 1880, Grattan-Guiness: 471-473). It was in 1823 and 1825 that Poisson proposed a particular method, a CI, and no less, obtained the Poisson integral formula from the Fourier development in series. This remarkable integral formula expresses the value of a harmonic function f at all points (r, φ) in polar coordinates, inside a circular disk of radius unity in terms of its values $f(1, \varphi)$ on the boundary of the disk:

$$f(r, \varphi) = \frac{1}{2\pi} \int_0^{2\pi} \frac{f(1, \psi)\left(1 - r^2\right)}{1 - 2r\cos(\varphi - \psi) + r^2} d\psi \tag{12}$$

where $r < 1$; the integral being possibly divergent in the limit $r \to 1$. After Fourier, CI began to be utilized as a powerful mathematical tool. A well-known example is the Dirichlet convolution kernel in Lejeune-Dirichlet's proof of the convergence of a Fourier series in (1829). In the same article, he also practically proved the commutativity properties of the convolution integral!

4 The Convolution and the Memory Problem

In the late 18th and the 19th centuries, the cooling of the earth became the object of intense study (Brush 1976: 551-566). For Fourier, the problem of the terrestrial temperature presented one of the most beautiful applications of the theory of heat (Fourier 1822: 3-4, 20). In (1821-1822) he published an astonishing article about terrestrial temperatures, and the diffusion of heat inside a spherical solid, subject to periodic temperature changes at its surface. Indeed the period was one year. For the first time, periodic boundary conditions appeared as a constraint with a partial differential equation. Fourier admitted that the examination of this problem would require multiple observations which were yet lacking. In 1825, he wrote a new memoir on the analytical theory of heat where he changed again the boundary conditions at the surface of a solid. Fourier proposed to study the effect when the temperature V at both extremities of a prism of length ℓ was subject to two time-dependent functions that may or may not be periodic. His mathematical model was a non-dimensional heat conduction equation:

$$\frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2}; \quad 0 < x < \ell; \quad t > 0, \tag{13}$$

with the boundary conditions:

$$\begin{aligned} V(0,t) &= \varphi(t), \\ V(\ell,t) &= f(t), \end{aligned} \tag{14}$$

and the initial value:

$$V(x,0) = \psi(x). \tag{15}$$

Here V is the temperature, t the time, x the distance from the first extremity; the distance from the second extremity will be represented by the variable ϖ.

With time-dependent boundary conditions, Fourier realized that if the time variable was made discrete, the solid did not have enough time to reach a stable state at each stage of discretization. Faced with this memory problem and an accumulation or a superposition of events, he therefore tried to represent a given value of a time-dependent boundary condition as a function of its past. Thus Fourier (1829) stated the solution to this question as follows:

$$\begin{aligned} V(x,t) \;=\; & \frac{x}{\varpi} f(t) + \frac{2}{\varpi}\sum_{i=1}^{+\infty}\frac{1}{i}\sin(ix)\cos(i\varpi)\left\{f(0) + \int_0^t f'(r)e^{-i^2(t-r)}dr\right\} \\ + \; & \frac{(\varpi - x)}{\varpi}\varphi(t) - \frac{2}{\varpi}\sum_{i=1}^{+\infty}\frac{1}{i}\sin(ix)\left\{\varphi(0) + \int_0^t \varphi'(r)e^{-i^2(t-r)}dr\right\} \\ + \; & \frac{2}{\varpi}\sum_{i=1}^{+\infty}\sin(ix)\int_0^{\varpi}\psi(r)\sin(ir)\,dr. \end{aligned} \tag{16}$$

Like the geometers of the 18th century, Fourier verified that his solution obeyed a non-dimensional heat conduction equation and the boundary conditions. He emphasized that the solution had three distinct parts. The first row represents the state where the point zero is restricted to having a zero temperature and the temperature of the other extremity varies as $f(t)$. Conversely in the second row, the extremity zero has a variable temperature $\varphi(t)$ and the other extremity is restricted to a zero temperature. In the third raw, the boundary conditions are both fixed at zero and the initial temperature of the solid has a distribution $\psi(x)$. In rows one and two, he obtains a solution in the form of convolution integrals $\int_0^t f'(r)e^{-t^2(t-r)}dr$ and $\int_0^t \varphi'(r)e^{-t^2(t-r)}dr$. Therefore, the memory effect and the superposition principle were established through these two convolution integrals. Note that the boundary conditions appear under the form of their derivatives $f'(r)$ and $\varphi'(r)$. This was due to the discretization and the quadrature of $f(r)$ and $\varphi(r)$.

Fourier was an extremely clever mathematical modeler. Because of the linearity of the partial differential equation he utilized the principle of superposition for the boundary conditions and the initial value, and thus was able to decompose his problem into smaller parts. Fourier's memoir was read at the Academy of Sciences in 1825, but not published until 1829, just one year before his death.

In the meantime, Poisson's memoir was read at the same Academy in 1826. His work concerned the theory of magnetism in motion (Poisson 1827). He was followed by Liouville (1832) and Duhamel (1833, 1834), particularly on the heat conduction equation with again time-dependent boundary conditions. Note that in the Anglo-German world, this principle of superposition (16) is known as the Boltzmann-Hopkinson principle (Boltzmann 1874, Hopkinson 1877).

After Duhamel, the work of French mathematicians seems to have been largely forgotten. Then in 1874 Boltzmann wrote *Zur Theorie der elastischen Nachwirkung* (the after effect of elasticity), a brilliant article on the superposition principle, memory effects, and convolutions. It is a very modern article, with a very different style from that of the French. He approaches the mechanical problem of torsion and the strain in an elastic bar as follows:

"(Boltzmann's) central hypothesis is that the deformation of a mechanical system is a function not only of the stress (i.e. of the forces) acting in a moment of the experiment but also of all the stresses acting on the system in all of its past" (Ianniello and Israel 2015).

By the end of the 19th century the convolution was accepted as a powerful mathematical tool, and the Italian mathematician Volterra (1887, 1913, 1924, 1928, 1931) wrote several books on the topic. Now, referring to the example of a simple Volterra integro-differential equation of convolution type:

$$x'(t) \quad = \quad x(t) + \int_{o}^{t} \varphi\,(t-s)\,x(s)ds + f(t). \tag{17}$$

This equation is said to be an equation with memory or sometimes a hereditary equation, because a convolution integral is included in the differential equation. Mathematicians began to acknowledge the importance of these types of equations near the end of the 19th century. Quoting Picard in 1907:

"*The examples are numerous where the future of a system seems to depend from anterior states: this is heredity. In such complex cases, it would be fitting to perhaps abandon differential equations and look towards functional equations, where would figure integrals taken from a very far-off time up to the actual time that will be the part of this heredity"* (Picard 1907).

5 Mathematics and Convolution Integrals

Towards the end of the 19th century, integral transforms, convolution products, and integro-differential equations all became the object of intense studies by Fredholm, Hilbert, and Volterra (see Kline 1972: 1052-1075). Volterra (1928, 1931) and others were also motivated by an interest in applications. Whereas, the convolution had been deducted from an intuitionist point of view in the 18th century, Volterra introduced his convolution integrals via an analogy with matrix products. In his

Princeton lectures, he was well aware of the commutative and associative properties of the convolution integrals (Volterra 1915).

In 1909, Lebesgue published an article on the existence of singular integrals $\int_a^b f(x)g(x)dx$. This integral exists if $f(x)$ is bounded and $g(x)$ is Lebesgue integrable or vice versa (Lebesgue 1909). Indeed, Lebesgue's theorem applies equally to convolution integrals.

Because the links between a convolution integral and a Laplace or Fourier transform are so important to their applications, we briefly present Borel's (1899) work on a "*Laplace like*" transform. Note that Mellin's work (1896) on that subject was unknown to Borel and the Paris circle at that time. Borel defined two functions $f(z)$ and $g(z)$ by their following "*Laplace*" integrals (Borel 1899, 1901):

$$f(z) = \int_0^{+\infty} F(u)e^{-u/z}\frac{du}{z}; \quad g(z) = \int_0^{+\infty} G(v)e^{-v/z}\frac{dv}{z}, \tag{18}$$

and then showed that the convolution integral is $H(x) = \int_0^x F(t)G(x-t)\,dt$. The Laplace transform of the convolution integral $H(x)$ reduced to a simple product of the two separate transforms $f(z)$ and $g(z)$. Borel failed to foresee all the possibilities of his theorem. Volterra, who maintained close relationships with the French mathematicians also did not see the possible uses of the theorem. But Borel's work turned out to be influential because in 1920, Doetsch produced a doctoral thesis at Göttingen University on Borel's summability theory of diverging series. Doetsch knew Borel's proof and was able to introduce modern, proper mathematical ideas on convolution integrals and Laplace transforms. The word *Faltung* was first introduced by Doetsch and Bernstein in 1920. The Laplace transform (Horn 1917, Doetsch, 1923) and the Fourier transform (Wiener 1933) are both adequate tools for evaluating a convolution integral. Note that Doetsch (1935) would introduce the convolution integral by analogy with a Cauchy product between two power series (Bradley and Sandifer 2009: ch 6). In his 1923 article, Doetsch proved the continuity of a convolution product, a fundamental property to smooth out any discontinuity. Then he proved rigorously that convolution is commutative ($f*g = g*f$) and associative ($f*(g*h) = (f*g)*h$.). Deakin (1982) would later remark that Doetsch surpassed all previous research in this area of study, but he was not without a dark side. Doetsch displayed a deep commitment to Nazis during Hitler's regime.

The convolution took on a new life with the Bourbaki group in France in the 1940s and 1950s, first with Weil's work (1940) on the integration in topological spaces and its applications, then with the theory of distributions (Schwartz 1951). Schwartz proved that the convolution product of two distributions T and S was also a distribution, which led to the identity $\delta * T = T$, where $\delta(t)$ is the Dirac distribution. With the commutativity and associativity properties of a convolution product, and taking $\delta(t)$ as a unity element, mathematicians chose to call this structure a pseudo-ring.

6 Some Applications

6.1 Theory of Probability

The theory of probability has implicitly generated many convolutions with a memory effect. We can take as demonstration the simple example of the sum k of the points of two dice $Z = X + Y$ to illustrate this. If the number of points of the first dice is $X = i$, the number of points for the second dice has to be $Y = k - i$, or vice versa and we obtain a triangular distribution function which is represented by a discrete convolution:

$$P(Z = k) = \sum_{i=1}^{k} p_X(i) p_Y(k - i), \tag{19}$$

where p_X and p_Y are the probability functions for the variables X and Y, respectively. The problem was solved for two or three dice by Cardano, Galileo Galilei, and Huygens. The probability function has already almost the shape of a Gaussian structure for three dice and illustrates the smoothing property and a spread of the distribution function known as the Central-Limit theorem in the theory of probability. At about the same time, in 1712-1713, de Moivre, James Bernoulli, and de Montmort solved the problem of n dice (Hald 1990: 36-41, 204-205).

The mystery of the Central-Limit theorem was fully solved in 1920. Concerned by the addition of independent random variables, Daniell (1920) showed that an equation of special importance occurred in connection with probability and other aspects of mathematical physics, namely:

$$f_X{}^* f_Y = f_{X+Y}. \tag{20}$$

In other words, the probability density function of the sum of two independent random variables f_{X+Y} is precisely the convolution product of the respective probability densities f_X and f_Y. In France, Levy (1925, 1937) completed the theory of the addition of n independent random variables with the additional illuminating properties of the convolution tool.

6.2 Filter Theory

Another application stems from the linearity properties of the convolution product and the fundamental theory of filters. The linearity of a convolution application seemed evident and of interest to engineers (Gardner and Barnes 1942), and these applications also possessed fundamental additive properties related to the principle of superposition. Now a system, for example, an electronic amplifier, is said to be invariant with respect to time if the same excitation $e(t)$, applied to a system $h(t)$ at different times, gives the same response $r(t)$:

$$\text{response system} = (\text{system function})^{*}\,(\text{excitation function})\,. \tag{21}$$

Such a system can be said to have no memory (Sirovich 1988: 80-81). Indeed, in any system, we may suspect that this behavior is valid only for a certain amount of time, and the Volterra's *heredity* may vanish hereafter. In Eq. 23, the system $h(t)$, a filter, for example, is often unknown. Already in the 1940s and during WWII in particular, engineers took an excitation $e(t)$ with the shape of a pulse for the calibration of a system. Because the Laplace transform of a pulse is unity, the Laplace transform of the response was the Laplace transform of the system (Gardner and Barnes, 1942: 255, 262-263).

6.3 Interpolation

In 1908, de la Vallée Poussin published *On the convergence of the interpolatory formulas between equidistant data*. His objective was first to re-examine the problem of equally sampled observations. The formula was (de la Vallée Poussin 1908, Godard 2000, 2004):

$$F(x) = \sum_{a}^{b} f\,(\alpha_k)\,\frac{\sin m\,(x-\alpha_k)}{m\,(x-\alpha_k)}, \tag{22}$$

where the α_k were the nodes, $f(\alpha_k)$ the ordinates, and $F(x)$ the interpolatory function. De la Vallée Poussin's intuition was simply based on the properties of $\sin c(x) = \frac{\sin x}{x}$; $\sin c(0) = 1$. With this interpolating kernel, the function $F(x)$ coincided with the observations. De la Vallée Poussin observed that his formula had clear application and was easy to implement without the instabilities linked to the Lagrangian polynomial. It was less sensitive to false points and, shared with the Fourier series, the fundamental property of point-wise convergence; the formula could be applied even if the function was only piece-wise continuous. VP's formula did not become a popular numerical tool, but it remains an elegant application of a convolution kernel.

6.4 Autoregression

Our final example is chosen from the approximation theory related to the time series analysis of sunspot activity. These dark patches on the surface of the sun appear periodically (Godard 1999), a periodicity in these numbers was first suspected by Horrebow in 1776. However, their study was complicated by random noise.

In the autoregressive model of Yule and Walker (Yule 1927, Walker 1931), the autoregression was essentially a regression related to past values of observations. In modern terms, an observation X_t at time t was replaced and smoothed by a weighted sum of past observations like so:

$$X_t = c_p + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t, \tag{23}$$

where c_p was a constant related to the order p of the regression, φ_i were the weights associated with each observation and ε_t was the random noise associated with the observation X_t. In this linear process, a window of observations was retained and the other values were erased. Yule and Walker thus realized that we should suppose *a certain degree of causal relationship between the successive observations*.

Acknowledgement The author would like to thank Mark Lewis for his help in editing this article.

References

Boltzmann L (1874) Zur Theorie der elastischen Nachwirkung. In: Wissenschaftliche Abhandlungen, Band (1865-1874): 616-644, Chelsea Publishing Company, New York, 1968

Borel E (1899) Mémoire sur les séries divergentes. Annales scientifiques de l'École Normale Supérieure, Paris, 16: 9-131

Borel E (1901) Leçons sur les séries divergentes. Gauthier-Villars, Paris

Bourbaki N (1960) Éléments d'histoire des mathématiques. Hermann, Paris

Bradley R, Sandifer E (2009) Cauchy's Cours d'analyse: an annotated translation. Springer, New York

Brush S G (1976) The kind of motion we call heat. North-Holland, Amsterdam

Clairaut A C (1743) Théorie de la figure de la terre: tirée des principes de l'hydrostatique. Available from the Gallica Internet site, Paris

Daniell P-J (1920) Stietjes-Volterra products. Congr. Intern. des Math., Strasbourg: 13-136

Deakin M A B (1982) The Development of the Laplace Transform, 1737-1937, II. Poincaré to Doetsch, 1880-1937. Archive for History of Exact Sciences, 26: 351-381

Deakin M A B (1981) The Development of the Laplace Transform, 1737-1937, From Euler to Spitzer, 1737-1880. Archive for History of Exact sciences, 25, 4: 343-390

Deakin M A B (1985) Euler's Invention of Integral Transforms. Archive for History of Exact Sciences, 33, 4: 307-319

de la Vallée Poussin C(1908) Sur la convergence des formules d'interpolation entre ordonnées équidistantes. Bulletin de l'Académie Royale des Sciences de Belgique, 1: 319-410

Dieudonné J (1981) History of functional analysis, North Holland, Amsterdam

Doetsch G (1923) Die Integrodifferentialgleichungen vom Faltungstypus. Math. Annalen, 89: 102-207

Doetsch G (1935) Der Faltungssatz in der Theorie der Laplace Transformation, Annili della Scuola Normale Superoire di Pisa, classe di Scienze 4, 1: 71-84

Dominguez-Torres A (2010) The origin and history of convolution I: continuous and discrete operations. Available from Internet

Dominguez-Torres, A (2015) A History of the Convolution Operation. IEEE Pulse, January 2015, available from Internet

Duhamel J M C (1833) Mémoire sur la méthode générale relative au mouvement de la chaleur dans les corps solides plongés dans des milieux dont la température varie avec le temps. J. École royale polyt., 14, 22: 20-77. Available from the Gallica Internet site
Duhamel J M C (1834) Mémoire Sur les Vibrations d'un système quelconque de points matériels. Journal de l'École royale polytechnique, XIV: 1-36, available from the Gallica Internet site.
Euler L (1763) Constructo aequationis differentiali $Aydu^2 + (B + Cu)dudy + (D + Eu + Fuu)$ $ddy = 0$, Sumto elemento du constante, Leonhardi Euleri Opera Omnia : I Volume 22: 395-402, E274
Fourier J-B (1818) Note relative aux vibrations des surfaces élastiques et au mouvement des ondes. Bulletin des Sciences par la Société philomathique: 129-136
Fourier J-B (1821-1822) Théorie de mouvement de la chaleur dans les corps solides. Mémoires Acad. Royale Sci. de Institut de France, t. V: 153-246, 1826
Fourier J-B (1822) The analytical theory of heat. English translation, Dover Publishers, New York, 1955
Fourier J-B (1829) Mémoire sur la théorie analytique de la chaleur. Mémoires Acad. Sci. Institut de France, t. VIII: 581-622
Gardner M F, Barnes J L (1942) Transients in linear systems studied by the Laplace transformation, Vol. 1: Lumped-constant systems, John Wiley and Sons, Inc, New York
Godard R (1999) An Historical Analysis of Time Series up to 1949. Historia Naturalis, 2: 305-328
Godard R (2000) The Interpolation Theory in the XXth Century. Proceedings of the CSHPM Annual Meeting, Hamilton, Vol. 13: 121-133
Godard R (2004) Sampling Theories from C. de la Vallée-Poussin to C. Shannon. Proceedings of the joint meeting CSHPM/BSHM, Cambridge, U.K., Vol. 17: 93-105
Godard R (2010) Some mathematical tools for numerical methods from 1805 to 1855. Proceedings of the 6^{th} ESU conference, Vienna
Grattan-Guiness I (1972) Joseph Fourier 1768-1830. The MIT Press, Cambridge, Massachusetts
Green G (1828) An Essay on the Application of Mathematical Analysis to the theories of Electricity and Magnetism. Originally published as book in Nottingham. Reprinted in three parts in Journal für die reine und angewandte Mathematik Vol. 39, 1 (1850): 73-89, Vol. 44, 4 (1852): 356-374 and Vol. 47, 3 (1854): 161-221
Hald A (1990) A history of probability & statistics and their applications before 1750. John Wiley & Sons, New York
Hopkinson J (1877) Residual charge of the Leyden Jar. Dielectric Properties of Different Glasses. Phil. Trans. R. Soc. Lond., 167: 599-626
Horn J (1917) Verallgemeinerte Laplacesche Integrale als Lösungen linearer und nicht-linearer Differentialgleichungen. Jahresb. Deutsche Math. Ver. 25: 301-325
Huygens C (1690) Treatise on Light. Translated in English by S. P. Thompson, reprinted by The University of Chicago Press, Illinois, 1950
Ianniello M G, Israe G (2015) Boltzmann's "Nachwirkung" and hereditary mechanics, available from Internet
Kline M (1972) Mathematical thought from ancient to modern times. Oxford University Press, Oxford
Lagrange J L (1773) Sur l'équation séculaire. Œuvres, Gauthier_Villars, Paris (1867-1892), Vol. 6: 349
Lagrange J L (1798) Mémoire sur les sphéroïdes elliptiques. Nouveaux Mémoires de Acad. Royale Sci. et Belles-Lettres de Berlin, années 1792-1793: 652
Laplace P S (1798-1823) Traité de mécanique céleste. Duprat J B M, Paris. Reprinted by Éditions Jacques Gabay (2006)
Laplace P S (1809) Mémoire sur divers points d'analyse. J. École poly.: 229-265
Lonseth A T (1977) Sources and applications of integral equations. SIAM Rev., 19, 2: 241-278
Lebesgue H (1909) Sur les intégrales singulières. in Œuvres scientifiques, Vol. III: 259-352, L'enseignement mathématique, Genève, 1972

Lejeune Dirichlet G (1829) Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données. Crelle, Journal für die reine und angewandte Mathematik, 4: 157-169

Levy P (1925) Calcul des probabilités. Gauthier-Villars, Paris

Levy P (1937) Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris

Liouville J (1832) Mémoire sur quelques questions de géométrie et de mécanique, et sur un nouveau genre de calculs pour résoudre ces équations. Journal de l'École Polytechnique, 13: 1-69

Mellin H J (1896) Über gewise duch bestimmte Integrale vermittle Bezichungen zwischen linearen Differentialgleichungen mit rationalen Coefficienten. Acta Societatis Scientiarum Pennicae, XXI, 6: 3-47

Picard E (1907) La mécanique classique et ses approximations successives. Révista di Scienza, 1: 4-15

Poisson S D (1823) Suite du mémoire sur les intégrales définies et sur la sommation de séries. Journal de l'École Polytechnique, : 404-509, available from the Gallica Internet site, and Mémoires de l'Académie de Sciences: 574, 1823 and Théorie de la chaleur

Poisson S D (1827) Mémoire sur la Théorie du magnétisme en mouvement. Mémoires de l'Académie des sciences de l'institut de France, 6: 441-570

Sachse A (1880) Essai historique sur la représentation d'une fonction arbitraire d'une seule variable par une série trigonométrique. Bulletin sci. math. et astro., 4, 1: 43-64

Sirovich L (1988) Introduction to Applied Mathematics. Springer-Verlag, New York

Schwartz L (1951) Théorie des distributions. tome II, Hermann & C^ie^, Paris

Volterra V (1887) Funzioni di linee. Scienze Matem, Roma

Volterra V (1913) Leçons sur les équations intégrales et les équations intégro-différentielles. Gauthier-Villars, Paris

Volterra V (1915) The theory of permutable functions. Princeton University Press, Princeton

Volterra V (1924) Leçons sur la composition et les fonctions permutables. Gauthier-Villars, Paris

Volterra V (1928) Sur la théorie mathématique des phénomènes héréditaires. J. Math. Pure Appl., 7: 249-298

Volterra V (1931) Leçons sur la théorie mathématique de la lutte pour la vie. Gauthier-Villars, Paris

Walker G (1931) On Periodicity in Series of Related Terms. Proceedings of the Royal Society of London, Ser. A, 131: 518-532. Gallica, 56224

Weil A (1940) L'intégration dans les espaces topologiques et ses applications. Hermann & C^ie^, Paris

Wiener N (1933) The Fourier integral & certain of its applications. Reprinted by Cambridge Mathematical Society, Cambridge, 1988

Yule U (1927) On a method of investigating Periodicities in Disturbed Series with Special Reference to Wolfer' Sunspot Numbers. Phil. Trans. (A) 226: 267-298

Sundials–Une Promenade Parisienne (A Math Walk in Paris)

George Heine

Abstract Sundials are among the many sites in Paris relevant to the history of mathematics. We describe a few of these sites and provide a method for estimating the accuracy of historic calendrical sundials.

Keywords Paris • Sundial • History

To those interested in the history of mathematics, the city of Paris offers numerous sites of interest. One can choose to interact with mathematics history in many ways, including:

- Reading names of mathematicians inscribed on historic buildings, including the Bibliotheque Ste. Geneviève and the Arc de Triomphe.
- Finding some of the many dozens of streets named after mathematicians; also the names of schools, parks, hotels, cafes, bars, and so forth.
- Finding pictures and statues of mathematicians, both in the museums and on the streets.
- Visiting the burial places of numerous mathematicians, in churches, in the Pantheon, and in the historic cemeteries. Mathematicians buried in Paris include Pascal, Descartes, Arago, Compte, Monge, Monge, Condorcet, Carnot, Painlevé, Lagrange, Fourier, Chasles, Compte, Jordan, and Poincaré.
- Locating places where significant events in mathematical history took place, such as
  - the location of the Couvent de Minimes, where Marin Mersenne held weekly meetings with leading scientists, mathematicians, and philosophers in the early seventeenth century;
  - the Pont de Neuilly, where Blaise Pascal experienced a near-death accident in 1658;

G. Heine (✉)
Independent Scholar, Pueblo, CO
e-mail: asteya@mathnmaps.com

M. Zack, D. Schlimm (eds.), *Research in History and Philosophy of Mathematics*, Proceedings of the Canadian Society for History and Philosophy of Mathematics
La Société Canadienne d'Histoire et de Philosophie des Mathématiques,
DOI 10.1007/978-3-319-64551-3_14

– sites of *salons*, where *savants* gathered to discuss science, mathematics, and philosophy during the Enlightenment;
– the site where Nicolas de Condorcet hid for several months in 1797 after being declared an Enemy of the State; and
– the site which housed the Ecole Polytechnique from 1805 until 1976.

In this brief article, we shall examine a few of the sites in the last category, specifically the locations of sundials. As a city founded in late antiquity, sundials were the chief means of time-keeping in Paris for many centuries. This history is commemorated in a street name in the 6th Arrondissement, the *Rue de Cherche-Midi*, "street of looking-for-the-south." The French word *Midi*, meaning "South," is a direct descendant of the ancient Latin word for south, *meri+dies*, literally "middle of the day." In the middle of the day, the azimuth of the sun, as seen from locations north of the tropics, is due south. Looking for the south is of course a necessary activity when constructing a sundial. The Latin *Meridies* is also the ancestor of both the English word "meridian" and its French cognate *meridien*.[1]

The original sundial in Rue de Cherche-Midi no longer exists, but there is a commemorative plaque southwest at 19, Rue de Cherche-Midi (Figure 1) with an old man showing an infant how to construct and use a sundial.[2]

Many buildings and streets in Paris feature old sundials. One attractive example (Figure 2) is at the entrance to the National Museum of the Middle Ages, the Musée de Cluny, in the 5th Arrondissement.

A different kind of sundial is found in one of the most visited sites in all of Paris, the Pyramid of the Louvre. In the room underneath the pyramid (the main

Fig. 1 Plaque at 19 Ru de Cherche-Midi

[1]There is a French colloquialism *chercher le midi a quatorze heures*, "to look for the south at 2 pm," meaning to complicate the issue, to look for a problem where there is none, to spend time and energy in a pointless endeavor.

[2]See (Hillairet 1993, Vol. 2, p. 183), for a history of the street, and p. 188 for a history of this plaque. Also see (Lalos 2010) for further information.

Fig. 2 Sundial at the entrance to the Musee de Cluny

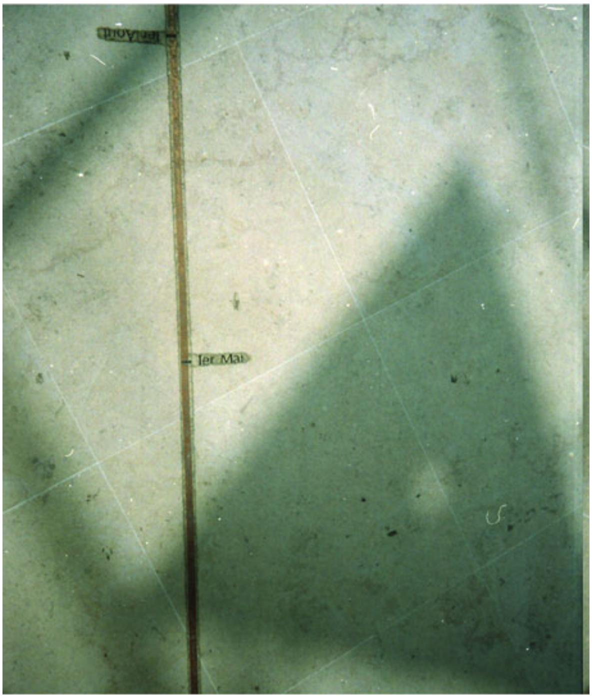


Fig. 3 Projection of the pyramid on the floor of the Louvre

entrance to the museum), there are marks in the floor showing the locations of the shadow of the pyramid at different times of the year – a *calendrical sundial*. The photo (Figure 3) of the floor of the Louvre shows the location of the shadow shortly

Fig. 4 Sundial at the Hotel de Invalides

before noon in late April 1999, close to a mark recording the position at noon on the first of May (1er *Mai*).

From the shadow, the gnomon appears to be a triangular plate with a circular hole, fixed to the frame of the pyramid, almost certainly on the south face.[3] At noon on the winter solstice, the altitude of the sun in the latitude of Paris is approximately 25°. The faces of the pyramid are sloped at about 33°, so a shadow near the top of the pyramid would lie outside its base, and thus not visible on the floor beneath for part of the year. It seems likely that the gnomon is located on the south face, and somewhat below the top.

Another calendrical sundial is found at the Hotel des Invalides, built by Louis XIV in the seventeenth century to care for wounded soldiers. This sundial is built to display both time and date. The attached photo (Figure 4) shows the location of the shadow on a morning in late April 1999.

In Paris, many calendrical sundials are found inside churches; for example, Figure 5 shows the interior of the Eglise de Saint-Sulpice. Before the era of clocks and calendars, these were useful in helping parishioners and clergy be aware of the coming of Easter.

A calendrical sundial displays dates based on the length of the sun's shadow, which is determined by the sun's elevation in the sky. Referring to the left part of Figure 6, the elevation of the north celestial pole is equal to the latitude ϕ. At the equinoxes, the sun (filled dot) is orthogonal to the polar axis, so its elevation at noon is equal to the complement of the latitude. At noon on the summer solstice, the elevation of the sun (open dot) is higher than at the equinox by an angle δ equal to the *obliquity*, which is the tilt of the earth's axis relative to the axis of its orbit.

The obliquity changes slowly over time, and it might be natural to inquire as to the accuracy of a calendrical sundial built several centuries ago. Today the obliquity

[3] The pyramid is not aligned with the cardinal directions; the south side actually faces south-southwest. However, the sun at noon shines most directly on the south side.

Fig. 5 Interior of Eglise de Saint-Sulpice

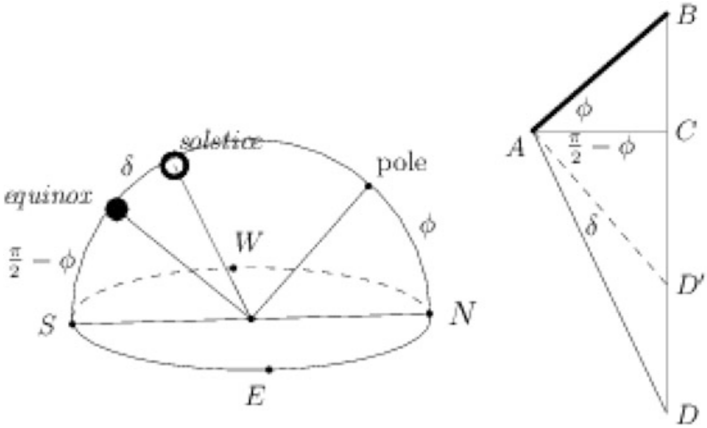


Fig. 6 Geometry of Calendrical Sundial

is about 23.44 degrees, while in the seventeenth century it was slightly larger, at about 23.48 degrees.[4]

[4]These figures were computed using Simon Newcomb's formula (see Various authors (2017)), considered very approximate today, but adequate for our purpose.

The right part of Figure 6 shows a typical design for a calendrical sundial.[5] The gnomon AB is affixed to a vertical wall BD, pointing north at an angle ϕ equal to the latitude. Thus, the gnomon is parallel to the earth's axis. If C is the point on the wall at the same height as A, then angle $BAC = \phi$. At noon on the equinoxes, the sun casts a shadow BD' exactly perpendicular to the earth's axis, so that angle $CAD' = 90° - \phi$, and the ratio $\frac{CD'}{CA} = \tan(90° - \phi)$. At noon on the summer solstice, the sun's elevation is increased by δ, so that the shadow is BD, and the ratio $\frac{CD}{CA} = \tan(90° - \phi + \delta)$. Taking $\phi = 48.855°$, the latitude of the Left Bank in Paris, today the value $\frac{CD}{CA}$ would be approximately $\tan(90° - 48.855° + 23.44°) \approx 2.1085$, while in 1675 the ratio would have been $\tan(90° - 48.855 + 23.48°) \approx 2.1045$. In other words, if the distance AC between the wall and the end of the gnomon were about 1 meter, the shadow would be about 4mm shorter today.[6] It seems likely that this would be smaller than the accuracy level of the original builders.

Paris offers the visitor many different glimpses of the history of mathematics. One could choose to look for names on signs, or to contemplate portraits of mathematicians. Or, as we hope to have shown in this brief article, one can find ways to interact with the mathematics itself.

References

Hillairet J (1993) *Connaissance du Vieux Paris.* In three volumes, bound together. Editions Payor et Rivages, Bd Saint-Germain, 75006, Paris.

Lalos M (2010) *Site de Cadrans Solaires.*http://michel.lalos.free.fr/cadrans_solaires; Cited 23 Apr 2017.

Various authors, 2017. *Axial Tilt* https://en.wikipedia.org/wiki/Axial_tilt; Cited 15 Jan 2017.

[5]This appears to be the design of the sundial on the Hotel des Invalides. It is *not* the design of the sundial on the Louvre pyramid, or in the Eglise de Saint-Sulpice, but similar arguments could be constructed for these.

[6]Metric measures are cited only to give the modern reader an idea of the scale of error; the meter, of course, was not yet invented in 1675.