UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
Confluence and Classification: Towards a Philosophy of Descriptive-Set-Theoretic Practice

Permalink
https://escholarship.org/uc/item/5md567n0

Author
Chen, Jason Zesheng

Publication Date
2024

Copyright Information
This work is made available under the terms of a Creative Commons Attribution-ShareAlike License, available at https://creativecommons.org/licenses/by-sa/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

Confluence and Classification: Towards a Philosophy of Descriptive-Set-Theoretic Practice

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Philosophy

by

Jason Zesheng Chen

Dissertation Committee:
Associate Professor Toby Meadows, Chair
Associate Professor Jeremy Heis
Associate Professor Martin Zeman

2024

© 2024 Jason Zesheng Chen

DEDICATION

Diese Dissertation ist den Proletariern aller Länder gewidmet.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Throughout my intellectual upbringing, I have been fortunate enough to have learned from lots of people smarter than me in lots of ways. The making of this dissertation would not have been possible without them.

My committee members, Toby Meadows, Jeremy Heis, and Martin Zeman, have taught me by example how to be a better philosopher, teacher, and mathematician. I have also benefitted immensely from conversations, mentorship, and guidance from various members in the philosophical and mathematical community, in particular Jeff Barrett, Matthew Foreman, Gabriel Goldberg, Joel David Hamkins, Juliette Kennedy, Shieva Kleinschmidt, Peter Koellner, Penelope Maddy and the Pen Pals group, Andrew Marks, Jeff Russell, Guozhen Shen, the Tobos group, Gabriel Uzquiano, Bokai Yao, Liang Yu, and audience at the UCI C-ALPHA Logic Seminar. I'd also like to thank the logic groups at Peking University and Fudan University, for inviting me over to basically rehearse my thesis materials in conference and various seminars.

I wish to especially thank Toby, whose advice to learn the Solovay model in my first year set the whole dissertation in motion. Since the beginning of each of these projects (and many others), Toby has been there to provide invaluable guidance and support through both the ups and downs of this journey. Thank you, for encouraging me to pursue whatever has interested me, for pushing me to do better, and for reeling me back in when I've gone astray.

I am grateful for my dear friends. First, Ephesus and Fat Choy: every time they jump on the keyboard a variable gets its name. Also Yuhao He, Weiwu Pang, Chao Wu, and Beinan Zhou, for being a delight for the past six years and for keeping my company with many jokes and conversations.

I must also thank and congratulate the newly minted Dr. Helen Shiyang Lu, my best friend and partner. One of the great pleasures of this Ph.D. is getting to do it alongside you. You show me how it's done (not that I'm anywhere near your level) and what it's like to be blessed with the presence of a better person.

I am beyond privileged to have been shielded from much of the slings and arrows of life. For this, and finally, I'd like to thank my family: Yong Chen, Yutuan He, and Zejing Chen. These loftier pursuits of mine are indebted to your honest toil. Thank you for your love, support, and vote of confidence.

VITA

Jason Zesheng Chen

EDUCATION

Doctor of Philosophy in Philosophy **2024**
University of California, Irvine *Irvine, California*

Master of Arts in Social Science **2022**
University of California, Irvine *Irvine, California*

Bachelor of Arts in Linguistics and Philosophy **2018**
University of Southern California *Los Angeles, California*

ABSTRACT OF THE DISSERTATION

Confluence and Classification: Towards a Philosophy of Descriptive-Set-Theoretic Practice

By

Jason Zesheng Chen

Doctor of Philosophy in Philosophy

University of California, Irvine, 2024

Associate Professor Toby Meadows, Chair

The dissertation presents a collection of interrelated works in the philosophy of mathematics. They are roughly unified by their focus on descriptive set theory, which is investigated through the lens of mathematical practice. Chapters 2 and 3 examine the roles that confluence plays in mathematical practice, such as providing justification for the Church-Turing Thesis. An extensive survey of the technical literature will attest to the ubiquity of justification by confluence, and it will be shown to serve a wide variety of justificational purposes that are largely orthogonal to each other. Chapter 4 presents a series of theorems and proofs that involve increasingly substantial use of metamathematical methods. Reflecting on the question of whether the metamathematical elements can be translated away without loss of insight, it attempts to shed light on our practical taxonomy of proofs by their methodology, as well as on the specific question of whether a proof can be said to make substantial use of metamathematical methods. Chapter 5 traces the pre-history of the theory of Borel equivalence relations, with the specific aim of identifying the early ancestors to this theory prior to its sudden emergence in the 1990s.

Chapter 1

Introductory Overview

1.1 Motivation

This dissertation consists of research works that are united in their focus on what mathematicians do, in particular descriptive set theorists and recursion theorists. Because each chapter is written in such a manner that they can be read alone (except chapters 2 and 3, which have a good amount of continuity), complete with their own motivations, introductions, and conclusions, let me say here a few words about the dissertation as a whole.

As the title suggests, I have chosen to look at descriptive set theory (mixed with some recursion theory) as my primary source of case studies. This is because descriptive set theory presents a fertile ground for interesting opportunities. As a proper mathematical subject, its fruitful interaction with other areas of mathematics lends itself to a rich trove of mathematical practices. At the same time, deep ties with metamathematics force it to remain tantalizingly close with the more foundational, extra-mathematical issues facing modern axiomatic set theory.

Like all works in philosophy of mathematics, the projects here will pay close attention to what the mathematics says; in writing this dissertation, however, I have also chosen to pay closer attention to what the mathematician says. Overall, the choice of topics in this dissertation is guided by a general position on the relation between philosophy and mathematics: "if our philosophical account of mathematics comes into conflict with successful mathematical practice, it is the philosophy that must give." (Maddy, 1997, p. 161) Rather than turning to this maxim as an arbiter of conflicting philosophies, I take this naturalist attitude as a rule of thumb for what kinds of questions to ask in studying the philosophy of mathematics: identify salient and successful mathematical practice first, figure out what purpose they serve and how they have come to be adopted, and then philosophize.

One upshot (or should I say trade-off?) of this approach is that the projects in here give off the impression that they are shying away from the admirable questions in the philosophy of mathematics, such as the nature of mathematical objects, the epistemology of mathematics, the status of axioms, or the ontology of mathematics. In exchange, however, I hope to provide a more detailed and nuanced account of what doing mathematics entails, as well as general frameworks in which mathematical activities beyond the stuff of proofs and theorems can be understood and analyzed.

With that said, let me now give a brief overview of the chapters in this dissertation.

1.2 Outline

Chapters 2 and 3 concern what I call justification by confluence, a salient pattern of justificatory practice across many mathematical disciplines. Its prototypical example: the Church-Turing Thesis, which states that any function intuitively computable by an effective method can be computed by a Turing machine. The received wisdom is that the thesis expresses the conviction that our formalization of effectivity by means of Turing machines is correct. This thesis is crucially justified by the equivalence of various definitions proposed in the 1930s, a phenomenon that I propose to understand in the more general context of justification by confluence.

Chapters 2 and 3 will focus on the roles confluence phenomena play in terms of evidence and justification. Via an extensive survey of the technical literature, I will show that this kind of justification turns out to be quite common across a diverse range of mathematical subfields, pervasive in actual mathematical practice. However, I shall argue that, despite an apparently singular application in the case of the Church-Turing Thesis, appeals to confluence actually serve a wide variety of justificational purposes that are largely orthogonal to each other.

Inspired by Maddy's poignant classification of what "foundational" means in mathematical/philosophical discourse (Maddy, 2019), I attempt to present a taxonomy of what these purposes are and show how they all arise, in some fuzzy and conflated fashion, in the case of computability. I shall also attempt to tease apart these roles in the surveyed examples, showing that they are in fact orthogonal to each other. At appropriate conjunctures throughout, I will also point to disagreements, debates, or open-ended philosophical questions in the literature, and apply this framework to show that they may be fruitfully resolved.

Turning more specifically to theorems and proofs, Chapter 4 works towards a two-fold goal: on the practical side, to provide a rough collection and classification of the various types of proofs in descriptive set theory that can be characterized by their use of metamathematical methods. In doing so, I also aim to track the degree of involvement of metamathematical tools in these proofs, and hope to demonstrate that there is a sui generis methodology, in full analogy with, say, "algebraic methods" or "topological methods". And on the philosophical side, the goal is to probe perhaps more general questions about the nature proofs and methods, particularly regarding what it means to characterize a proof by the method used, and what a method imports in a proof.

To do this, I will first motivate this project a little by relating it to a broader philosophical project born out of Dawson's monograph (Dawson Jr., 2015) about the values of having multiple proofs. Next, I will consider an obvious non-example of metamathematical proof (that makes use of metamathematical methods), which will allow me to sketch a list of putative objections to calling such proofs metamathematical. These putative objections will shed light on the practice of organizing mathematical proofs by their methodology. Following that, I will present a series of theorems and proofs that involve increasingly substantial use of metamathematical methods. At each turn, I will analyze how the proof in question addresses the earlier objections, honing in on the issue of whether the metamathematical methods can be translated away without loss of insight. This sort of dialectics will hopefully shed light on our practical taxonomy of proofs by their methodology, as well as on the specific question of whether a proof can be said to make substantial use of metamathematical methods.

Chapter 5, the final chapter, traces a significant but under-addressed piece of the history of descriptive set theory. The common story of the history of descriptive set theory tells of Cantor's ingenious venture into the transfinite, how his work found palpable expression with the French analysts, the ill-fated program of the Moscow school to understand the projective hierarchy, and its marvelous revival in the west with large cardinals and determinacy. Nevertheless, a cursory glance at some of the recent publications in generalist mathematics journal such as the Annals of Mathematics and Inventiones Mathematicae reveals that the field has been making remarkable contributions to a wide range of mathematical areas, including ergodic theory, functional analysis, measure theory, and graph combinatorics, with techniques and concerns that have little motivation in ways of classical concerns such as the continuum hypothesis or large cardinals.

This of course is due to the rich theory, developed in the past 30 years, of invariant descriptive set theory (or the theory of Borel or otherwise definable equivalence relations).

And yet the sudden emergence of the theory in the 1990s appears mysterious. Chapter 5 is a partial attempt to dispel this air of mystery. Doing this will involve some folklore archaeology, so to speak. Granted, the mathematical contexts immediately preceding the birth of the theory of Borel equivalence relations are well-known to experts in the field and are scattered in the introduction sections in various specialist literature. These include works in disparate fields of mathematics, somewhat distant from descriptive set theory, that were ultimately understood and absorbed in the context of (non-)reducibility theorems in invariant descriptive set theory. In light of this, one might naturally wonder if any such precursors can be found in descriptive set theory proper. In Chapter 5, I identify a number of theorems, techniques, and programs that later figure in the emergence and development of invariant descriptive set theory, pre-dating the materials commonly found in the relevant specialist literature. The goal is to show that certain central concerns of invariant descriptive set theory has been at the heart of the early developments of descriptive set theory itself.

Chapter 2

Justification by Confluence, Part I: Stuff of Practice

2.1 Introduction

The Church-Turing Thesis states that any function intuitively computable by an effective method can be computed by a Turing machine. The received wisdom is that the thesis expresses the conviction that our formalization of effectivity by means of Turing machines is *correct*. Aside from Turing's insightful analysis of the meaning of *effective method*, the thesis is crucially justified by a certain phenomenon: all attempts to formalize the notion of effective method have led to the same class of functions.

I propose to understand this last phenomenon in the more general context of justification by confluence. Roughly put, confluence[1] refers to situations whereby a number of distinct attempts (to define, formalize, solve, etc.) have all produced results that are in some sense

[1]To borrow a term from Gandy (1988). Gandy's notion of confluence has recently been significantly elaborated by Kennedy (2013, 2020), especially in relation to entanglement with formalisms. Here, I will use the term much more liberally, to encompass a wider variety of phenomena.

equivalent or conversely that a particular object or concept admits numerous seemingly distant equivalent definitions. In particular, Chapters 2 and 3 will focus on the roles confluence phenomena play in terms of evidence and justification.

Of course, since as early as Gödel's *What is Cantor's Continuum Problem?*, justification has been a staple in modern philosophy of mathematics. One readily finds a rich trove of existing work: justification of axioms (Maddy, 1988a, 1988b), justification of mathematical belief and knowledge (Clarke-Doane, 2015), justification of certain practices (Antos, 2022), what forms justification can take (Barton et al., 2020; Heron, 2021), etc. So to add anything substantial to the existing literature on justification, one should at least be clear on what is being justified and how it is justified.

As a first pass, the prototypical example of justification by confluence manifests itself in the justification of the Church-Turing Thesis. The homonymous entry in the Stanford Encyclopedia of Philosophy (Copeland, 2024) cites the following remark by Church:

> The fact ... that two such widely different and (in the opinion of the author) equally natural definitions of effective calculability ... turn out to be equivalent adds to the strength of the reasons adduced below for believing that they constitute as general a characterization of this notion as is consistent with the usual intuitive understanding of it. (Church, 1936)

The seminal textbook by Kleene (1952) adds many other equivalences to the list. Jointly, the growing list of equivalences is meant to convince one of the truth of the Church-Turing Thesis. Kleene cites this phenomenon, among others, as evidence for Church's Thesis:

> Several other characterizations of a class of effectively calculable functions ... have turned out to be equivalent to general recursiveness, i.e. the classes of functions which they describe are coextensive. ... The fact that several notions which differ

> widely lead to the same class of functions is a strong indication that this class is fundamental. (Kleene, 1952, p. 319)

Following Gandy, we will occasionally refer to this line of thinking in justifying the Church-Turing Thesis as "(Church's or Kleene's) argument by confluence". Gödel, in his 1946 Princeton Bicentennial Address (see Kennedy (2013)), similarly attributes the success of Turing's analysis of computability to this confluence: "It seems to me that this importance [of Turing computability] is largely due to the fact that with this concept one has succeeded in giving an absolute definition of an interesting epistemological notion, i.e., one not depending on the formalism chosen." Our modern vantage point has only made this argument more forceful for Copeland:

> All attempts to give an exact characterization of the intuitive notion of an effectively calculable function have turned out to be equivalent, in the sense that each characterization offered has been proved to pick out the same class of functions, namely those that are computable by Turing machine. The equivalence argument is often considered to be very strong evidence for the thesis, because of the diversity of the various formal characterizations involved. (Copeland, 2024)

To echo this appeal to diversity and confluence, and to point out that this mode of justification extends beyond just the Church-Turing Thesis, let us also take a hint from Maddy's classic paper on axiom justification (Maddy, 1988a). There, surveying the various rules of thumb by which set-theoretic axioms came to be accepted, she notes Kanamori and Magidor's remarks (Kanamori & Magidor, 1978) that the many equivalent definitions of weakly compact cardinals serve as "good evidence for the naturalness and efficacy of the concept". Recalling that the same pattern of argument was observed in early discussions of the naturalness of the concept of general recursiveness, Maddy writes:

> ... [T]he property of weakly [sic] compactness is equivalent to the compactness of the language $L_{\kappa,\kappa}$, and to a certain tree property, and to an indescribability property, and to several other natural properties ... This convergence has led some writers to *diversity*, another rule of thumb. (Maddy, 1988a, p. 504)

Later in the same paper, Maddy highlights another instance of appeal to this rule of thumb by Kanamori and Magidor, this time on the subject of zero-sharp and the list of statements equivalent to its existence:

> ... [The assumption that zero-sharp exists] turns out to be equivalent to a determinacy assumption and to an elementary embedding assumption. This prompts Kanamori and Magidor to another application of diversity: '[This] is a list of equivalences, much deeper than the confluence seen at weak compactness.' Thus the implication of the existence of the sharps provides a very appealing extrinsic support for the Axiom of Measurable Cardinals. (ibid., p. 507)

It's fair to say that in each of the preceding examples, something is being justified, although it is decreasingly clear what is being justified and how. It seems to me that the case of the Church-Turing thesis is relatively clear. The definitions were intended for achieving a specific common goal: to capture the informal notion of effective feasibility (whatever that may be). This goal had sufficiently significant motivation: both Hilbert's tenth problem and his Entscheidungsproblem turned on its satisfactory resolution. And the common lesson drawn from the equivalence between, say, the λ-definable functions and the Turing-computable ones is that we have the *correct* formalization of the informal notion of effective/mechanical computability. Against this backdrop, the Church-Turing thesis has a clear philosophical role to play: it is the expression of (perhaps our confidence in) the correctness of our formal conceptual analysis of idealized effective feasibility.

On the other hand, the other two cases apparently lack many of the above factors that made the Church-Turing Thesis so appealing and its attendant evidence interesting. For Kanamori and Magidor, the many equivalent definitions of weak compactness was considered "good evidence for the naturalness and efficacy of the concept". Other than this, one can say very little by way of the kind of justification seen the arguments for the Church-Turing thesis.

For example, neither the tree property nor the generalized Ramsey-theoretic property was designed to capture some nebulous, pre-theoretic notion that ended up being formalized by weak compactness. There is no single notion that motivated the various definitions of weak compactness; quite on the opposite, the various equivalent definitions arose independently via generalizations from finite to the infinite[2]. Whatever else the confluence of these definitions may have justified, correctness of formalization ranks very low on the list, if at all.

Even if one succeeds in coming up with a plausible candidate notion that weak compactness is supposed to have captured, the situation zero sharp is even more puzzling[3]. No doubt the equivalence between the existence of zero sharp (a small combinatorial object) and the existence of an elementary embedding $j : L \to L$ is significant. Such equivalences underlies much of the works in early inner model theory and even guided the appropriate definitions for generalized sharps. But what exactly does the equivalence tell us, aside from providing us with diverse tools to study it? And for that matter, what is it about this equivalence that makes it "deeper than the confluence seen at weak compactness"?

I hope the brief discussion above illustrates what makes justification by confluence intriguing. Indeed, there is something salient about confluence that intuitively imports justification. An initial goal of this work is to figure out what is being justified and how. Taking Kleene's

[2]And indeed, the original application of these compactness cardinals was to study the size of measurable cardinals, by giving historically the first proof that there are many inaccessible (and much more) cardinals below the least measurable cardinal, a result that was "greeted as a spectacular success for metamathematical methods" (Kanamori, 1996). Also see Chapter 4 of this thesis for a study of metamathematical methods.

[3]Even more puzzling still, considering the folklore that this line of work was born out of Jack Silver's attempt to prove the inconsistency of measurable cardinals.

argument by confluence as a prototypical instance, and abstracting away, our target of analysis manifests in the following speech pattern found in the practices of mathematicians, logicians, and philosophers.

(*Conf*) The fact that we have all these equivalences or otherwise similar results constitutes evidence for...

It turns out statements of this form are quite common across a diverse range of literature, not just in computability and large cardinals. Much of this chapter and the next will be dedicated to analyses and discussions of them. Hence their subject matter, and what I consider to be its main original contribution, is the isolation of justification by confluence as a salient pattern of justification. Via an extensive survey of the technical literature, I will show that this pattern is pervasive in actual mathematical practice, ranging from computational complexity, higher recursion theory, and descriptive set theory, to algorithmic randomness, Borel equivalence relations, and dynamical systems.

At the same time, granting that certain confluence results do seem to provide nontrivial justification for certain beliefs or practices, I will look at a vast array of situations in which confluence results have led mathematicians to express something analogous to the Church-Turing Thesis in spirit. And so I aim to (at least partially) survey of the various instances where this kind of language occurs in practice and to present an analysis of what is the underlying justificatory process behind them. In this context, I will be primarily concerned with what kinds of statements can take the place of the ellipsis in (*Conf*) and what roles confluence plays in their justification. My goal in doing so is to classify the kinds of practical roles they play in the context in which they are uttered.

The practical contribution of this work is to articulate a taxonomical framework, through which appeals to confluence (as well as claims about theses of the Church-Turing type) can be better understood. Familiar readers will surely notice that this is heavily inspired by

Maddy's poignant classification (Maddy, 2019) of what "foundational" means in mathematical/philosophical discourse comparing the roles played by set theory, category theory, and homotopy type theory. As such, this work is going to be as much about the nature of various instances of confluence as it is about the actual mathematical practice of invoking them to justify one thing or another.

More specifically, I shall demonstrate that, despite an apparently singular application in the case of the Church-Turing Thesis, appeals to confluence actually serve a wide variety of justificational purposes that are largely orthogonal to each other. We shall see, for example, that the justificatory potential for these confluence results does not necessarily hinge on a pre-existing informal notion to be captured: there are a number of appeals to the Church-Turing Thesis that are really appeals to confluence in disguise; and there are a number of instances of (*Conf*) in which confluence gestures at something other than successful formalization. I will try to convince the reader that, once this perspective is adopted, we actually observe, even in the most quotidian practices of mathematics, many diverse justificational purposes when confluence is invoked. As a rough overview, these justificational purposes can be classified as follows:

1. **Conjecture Heuristic**: in this form with the least philosophical baggage, confluence motivates conjectures of the form "every such-and-such will be so-and-so", on the basis that a large class of (natural, practical) examples are so-and-so. A typical example is Martin's Conjecture in recursion theory.

2. **Rigor Assurance**: in terms of theorems and proofs, confluence allows mathematicians to be confident that formal rigor is not loss in translation between informal and formal language or translation between language used in different fields of mathematics. This allows for freedom to move at once between different levels of rigor and conceptual frameworks.

3. **Coding Invariance**: confluence suggests that our theorems are not a result of superficial coding peculiarities. The same result is obtained regardless of the coding of the objects/concepts, indicating that we have gained bona fide knowledge about the objects/concepts themselves rather than being misled by formalisms.

4. **Joint-carving**: this is the received wisdom of the Church-Turing Thesis, that a certain attempt to precisify/formalize an informal idea has successfully "carved nature at its joints", i.e., correctly identified the right class of objects instantiating a natural concept. Generalizing a little, this is the idea that confluence suggests that something has been done correctly. In particular, on a commonly accepted *no-accident* interpretation of such phenomena, attempts to analyze such curiosity have been a main motivator for certain scientific/mathematical investigations.

5. **Remarkable Coincidence**: in contrast, in this case confluence suggests that something has been done fruitfully. That is, confluence facilitates further investigations, providing tools and insights without presupposing that they instantiate any particular notion.

6. **Counterexample Resistance**: confluence reveals a robust underlying notion, one that is not easily refuted by counterexamples and immune to various sterngthenings and weakenings.

Among these, the first three are more practice-oriented and are treated in this chapter. The latter three contain philosophical claims that are more involved and are relegated to Chapter 3. Together, Chapters 2 and 3 will be structured accordingly, dedicating each section to a particular boldfaced item above. It will be seen that all of the above arise, in some fuzzy and conflated fashion, in the case of the Church-Turing Thesis. I shall also attempt to tease apart these roles in the surveyed examples, showing that they are in fact orthogonal to each other. To do this, each section will begin with observed examples of the titular justificational role played by confluence in the context of computability and the

Church-Turing Thesis, followed by further, more foreign examples to be analyzed, showing that they appeal perhaps exclusively to one particular role and not others.

At appropriate junctures throughout this chapter and the next, I will also point to disagreements, debates, or open-ended philosophical questions in the literature, and I will show how they may be fruitfully subsumed under the framework developed here. These include San Mauro's diagnosis of the tension between what he calls the practical side of the Church-Turing Thesis and the "Standard View" about its relevance; Hamkins's call for a philosophical justification of what he calls Gao's Thesis in the study of Borel equivalence relations; the putative status of a "Church-Turing Thesis" for algorithmic randomness; and finally the significance about recent confluence results regarding constructible hierarchies relative to generalized logics.

2.2 Conjecture Heuristic

Let us begin with something that carries the least philosophical baggage[4]. In his recollection of the early days of λ-calculus, Kleene recalls coming up with the definition of the predecessor function on the natural numbers[5] in 1932 (at the dentist's office, of all places) and showing the result to Church, who "had just about convinced himself there wasn't a predecessor function [definable in λ-calculus]" (Crossley, 2006, p. 5). This and other related results inspired great confidence in the range of functions that could be defined in λ-calculus. Church, for example, gradually came around and asked "whether [they] had not really got *all* the effectively calculable functions."

At any rate, that eventually became the prevailing sentiment at Princeton:

[4]Least for the present context, anyway. I concede that probabilistic thinking and conjecturing in mathematical practice remains a tricky topic in the philosophy literature.

[5]The motivation arose out of Kleene's approach to deriving the Peano axioms in Church's original formalism (which was eventually shown inconsistent by Kleene and Rosser), in particular the injectivity of successor. See Kleene (1981, p. 56) for the history.

> There was no idea [before the discovery of the λ-definable predecessor function] that [the class of λ-definable functions] was going to be all effectively calculable functions. But I [Kleene] kept taking it as a challenge and everything I tried I could work [sic], and [having been able to define the integers] we got the idea that this could represent all calculable functions ... for us the first idea that λ-definability was general was after ... discovering that everything you thought of that you wanted to prove λ-definable, you could. (Crossley, 2006, pp. 5–6)

Keeping in mind that this development took place in what can be said to be the pre-historic era of the Church-Turing Thesis, prior to any knowledge of Turing's work in 1936, the confluence at play here is a particularly simple one: all the effectively calculable functions one could think of were λ-definable. And this seemed to have increased the credence that the latter captures all of them.

To be fair, the quoted passage still has quite a bit in common with the later Church-Turing Thesis. For one, it seems to express that they had done something right in their formalization of effective calculability. However, I would like to entertain the possibility that something else was conveyed alongside this vindication. To see what, let me quote from a letter from Church to Kleene, dated November 29, 1935, according to which Gödel regarded Church's proposal of identifying effective calculability with λ-definability as "thoroughly unsatisfactory" (Davis, 1982, p. 9), presumably sometime around his visit to the Institute for Advanced Study in the fall of 1933. Despite this, Church recalled remaining confident: "If [Gödel] would propose any definition of effective calculability which seemed even partially satisfactory, [Church] would undertake to prove that it was included in lambda-definability [sic]." (ibid.)

Gödel did in fact propose a definition of effective calculability. In a series of lectures at the IAS[6] in 1934, Gödel presented what is known today as the Herbrand-Gödel general recursive

[6] Transcribed in Gödel (1934) and reprinted in Gödel (2001, pp. 338–371).

functions. Kleene recounts questions being asked along the lines of "Does this embrace all effectively calculable functions, and is it equivalent to λ-definability?" (Crossley, 2006, p. 6) For us, this draws out a useful distinction: suspending judgment on the correctness of one's own definition, one can still independently conjecture that it is equivalent to another plausible definition.

Even though Gödel at the time[7] was "not at all convinced that [his] concept of recursion comprises all possible recursions" (Davis, 1982, p. 8), Church, who had "explicitly come out with [Church's Thesis]" (Crossley, 2006, p. 6), responded by proving the equivalence between the two notions together with Kleene (according to whom, "then it was a simple matter to prove the equivalence of the two notions" (ibid.)).

Here lies a plain justificatory role played by confluence: along with the conviction that λ-definable functions are the correct formalization of effective calculability, Church was also justified in conjecturing that other plausible definitions will be (either weaker or) equivalent[8]. This seems straightforward enough: if λ-definability is the right definition for the effective calculabiility, then any other candidate definition worth their salt should be equivalent to it. The point is that, after seeing just the mini-confluence (all the functions they could think of were λ-definable), Church gained not only credence in the correctness of his definition, but also a heuristic for conjecturing that other definitions will be equivalent to it.

This kind of **Conjecture Heuristic** was first documented in print by Post (1936), when he tried to provide yet another definition of effective calculability (now known as the Post-Turing machines, one that he considered to be of "logical potency" and "psychological fidelity"). Acknowledging the earlier equivalence between Church's λ-definable functions and

[7]For an evolution of Gödel's view on computability, see Sieg (2006). Gödel would go on to be convinced by Turing's 1936 analysis. We will return to this in later sections.

[8]In an abstract of Church (1936) submitted to the *Bulletin of AMS* in 1935, Church wrote: "... it is maintained that the notion of an effectively calculable function ... should be identified with that of a [general] recursive function, since other plausible definitions of effective calculability turn out to yield notions which are either equivalent to or weaker than [general] recursiveness." (Davis, 1982, p. 10)

Herbrand-Gödel general recursive functions, Post ends his paper by speculating that his machines are also equivalent: "The writer expects the present formulation to turn out to be logically equivalent to recursiveness in the sense of the Gödel-Church development."

Of course, the modern-day reader will know that virtually all attempts to define effective calculability have ended up equivalent. With the Church-Turing Thesis fully fledged, we see the same kind of conjecture heuristic at work when Wolfram conjectured that a particular class of cellular automata can simulate any Turing machines. The SEP entry *Cellular Automata*[9] documents one such conjecture:

> The basic feature a cellular automaton needs to perform computations is the capacity of its transition rule of producing ...localized, stable, but non-periodic configurations of groups of cells ... seen as encoding packets of information ... [which] can propagate in time and space without undergoing important decay. The amount of unpredictability in the behavior of Class_4 rules also hints at computationally interesting features: by the Halting Theorem ... it is a key feature of universal computation that one cannot in principle predict whether a given computation will halt given a certain input. These insights led Wolfram to conjecture that Class_4 [cellular automata] were ... capable of universal computation. ... Rule 110 [a particular cellular automaton] was indeed proved to be computationally universal. (Berto & Tagliabue, 2023)

The preceding example is meant to further tease apart the two types of sentiment somewhat present in Church's letter. Church was confident on the one hand that his definition was correct, and on the other hand this warrants the conjecture that other similar definitions would be equivalent. Wolfram's conjecture, however, does not seem to presuppose one way or another whether he had any correct definition of effective calculability. That is, having the

[9] To which I refer the reader for a survey of the requisite background.

history of confluence of computability notions in the background and the precedent that certain cellular automata can indeed simulate universal Turing machines (e.g., Conway's Game of Life), one is already rationally licensed to conjecture that a celluar automaton equipped with rules that make it resemble effective computation is Turing-complete.

This heuristic for conjecturing that certain rules of cellular automaton make it able to simulate (perhaps all) Turing machine computations is exactly as Rogers spelled out in his seminal textbook on computability:

> In fact, if certain general (and reasonable) formal criteria are laid down for what may constitute a [specification of what counts as computation], it is possible to show that the class of partial functions obtained is always a subclass of the maximal class of all partial recursive functions. (Rogers, 1987, p. 18)

The point I want to make is a simple one. If one sees a number of (natural) attempts or instances have all converged on a particular result, then one is licensed to conjecture that the next one is going to yield the same result, or even better, that all such attempts/instances will end up being in some sense equivalent. This reasoning nothing above and beyond just ordinary empirical induction. Indeed, such a heuristic is so common that it is almost paradigmatic of much of the later work in computability theory. To point to a prominent example, the most famous open problem in computability theory[10][11]:

Conjecture (Martin's Conjecture)**.** *Every Turing-invariant Borel function is either constant on a cone or Turing-equivalent to an iterate (maybe transfinite) of the Turing jump on a cone.*

[10] I have chosen to present a lighter version aimed at general mathematical audience, taken from Montalbán (2019), which differs from the stronger and more general version found in the specialist literature. The reason is that the latter contains more set-theoretic jargons and subtleties that will distract us from the current exposition, with little gain in the lesson.

[11] A function $F : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ is Turing-invariant if and only if it maps Turing-equivalent reals to Turing-equivalent reals, i.e., $x \equiv_T y$ implies $F(x) \equiv_T F(y)$. For example, the Turing jump is Turing-invariant. A cone is a set of the form $\mathrm{Cone}(a) := \{y \mid a \leq_T y\}$. We say F is constant on a cone if there is a cone whose image under F contains only one Turing degree.

The heuristic for this conjecture partially follows from an empirical observation. Having surveyed a number of natural examples of uncomputable sets, Montalbán (2019) summarizes:

> [The examples ordered under Turing-reduction] do seem to form a hierarchy. There are many more examples one can get from elsewhere in mathematics and many more from computability theory that are still very natural ... All the examples we know are ordered in a line. Furthermore, we know of no natural example strictly in between 0 and K [the halting problem], and all natural examples strictly in between 0 and COF [the set of polynomials in $\mathbb{Z}[x, y_0, y_1, y_2, ...]$ that have integer solutions for all but at most finitely many values of x] are Turing equivalent to either K or TF [finite presentations of torsion-free groups]. We know many natural examples between COF and TA [truth of arithmetic], though they are nicely ordered in a line like the ordering of the natural numbers. We know many natural examples between TA and WF [codes for programs computing well-founded trees], and they are nicely ordered in a well-ordered line.

Here one hears echoes of confluence: *all the natural examples of uncomputable sets end up being linearly ordered by Turing-reduction.* Considering the diverse fields that these examples come from, the following observation of further confluence surely reminds one of Kanamori and Magidor's appeal to *diversity*:

> The halting problem K has degree $0'$, and one can show that so do WP [word problem for finitely generated groups] and HTP [Hilbert's tenth problem]. And if we apply the Turing jump again, we get $0''$. It turns out that TF is equivalent to $0''$. If we take another jump, we get to $0'''$, which happens to be Turing equivalent to COF (ibid.).

All the natural examples of uncomputable sets end up being linearly ordered. All of them are Turing-equivalent to iterates of Turing jumps. This so-called "linearity phenomenon" marks yet another example of confluence that has led to a conjecture. According to Montalban, "Martin's conjecture is a formal statement trying to capture the essence of ... the following empirical observation: while the infinite sequences in $\mathbb{N}^{\mathbb{N}}$ are not linearly ordered by Turing computability, the naturally occurring sequences are." (ibid.)

Conjectures similar in spirit to Martin's conjecture abound in mathematical logic. The confluence engendered by the linearity phenomenon[12] has a particularly close relative in axiomatic theories[13]. The abstract of S.-D. Friedman et al. (2013) begins with the proclamation: "The fact that 'natural' theories, i.e. theories which have something like an 'idea' to them, are almost always linearly ordered with regard to [consistency] strength has been called one of the great mysteries of the foundation of mathematics." This observation led Montalban and Walsh to prove an analogue in Montalbán and Walsh (2019)[14].

Elsewhere, Bagaria (2023) has shown that a variety of large cardinal principles are equivalent to instances of what he calls the general principle of Structural Reflection[15]. For Bagaria, the equivalences serve to induce confidence in the assertion that "all large cardinals are in fact different manifestations of a single general reflection principle." And eventually this led to the following conjecture in Bagaria and Ternullo (2023): "Conjecture 1. every known large cardinal notion is equivalent to some (form of) [structural reflection]."

[12] For a proposed counterpoint, see Hamkins (2022).

[13] There is yet another close relative, i.e., all natural theories extending ZF are equiconsistent with large cardinal axioms and well-ordered by consistency strength. This phenomenon has not motivated overarching conjectures like Martin's conjecture, but rather piecemeal consensus, e.g., that the determinacy hypotheses would be equiconsistent with some large cardinals (which we now know) or that Proper Forcing Axiom will be equiconsistent with a supercompact cardinal (which we don't). So I would like to consider this phenomenon as serving some other purpose that we will return to later, rather than guiding conjectures, although it certainly does so at points.

[14] I refer the interested reader to recent works by James Walsh, in particular Walsh (2024) where both the philosophy and the mathematics are elaborated.

[15] This refers to principles of the form "For every definable, in the first-order language of set theory, possibly with parameters, class $\mathcal{C}$ of relational structures of the same type there exists an ordinal α that reflects $\mathcal{C}$, i.e., for every A in $\mathcal{C}$ there exist B in $\mathcal{C} \cap V_\alpha$ and an elementary embedding from B into A." (Bagaria, 2023, p. 30)

Recall that Church was confident that whatever definition Gödel proposed would be equivalent to his own. That confidence stemmed partly from his belief that he had the right definition. In contrast, Bagaria supplies a disclaimer to the effect that he "[does] not wish to claim that the [structural reflection] principle is what Godel had in mind when talking about reflection of internal structural properties of the membership relation." So for Bagaria, structural reflection principles are not intended as an attempt to capture any pre-theoretic notion, but rather they are an interesting family of principles, motivated from what Cantor and Ackermann had said about absolute infinity, to which many large cardinal notions happen to be equivalent. It is this "complex, and multi-faceted, ramification of principles, and the equivalence results with large cardinal notions obtained" that led him to conjecture that all large cardinals are equivalent to some form of structural reflection.

For a final example of **Conjecture Heuristic** provided by confluence, let us turn to another most famous open problem: P vs. NP. In the survey Aaronson (2016), Aaronson makes numerous cases for conjecturing $P \neq NP$. In doing so, he managed to sum up, in my view, what it is it that provides **Conjecture Heuristic** from confluence:

> To my mind, however, the strongest argument for $P \neq NP$ involves the thousands of problems that have been shown to be NP-complete, and the thousands of other problems that have been shown to be in P. If just one of these problems had turned out to be both NP-complete and in P, that would've immediately implied $P = NP$. Thus, we could argue, the hypothesis has had thousands of chances to be "falsified by observation." (Aaronson, 2016, p. 25)

Certainly, here Aaronson is telling the usual story of inductive confirmation that we are adequately used to in the empirical sciences. But the added strength from diversity makes it resemble Kleene's argument by confluence. That is, for all the discoveries of NP-complete and P-problems, all from diverse areas, no one has been able to equate the two classes[16]. Aaronson (2014) elaborated on this "invisible electric fence" with a meticulous list:

> By now, tens of thousands of problems have been proved to be NP-complete. They range in character from theorem proving to graph coloring to airline scheduling to bin packing to protein folding to auction pricing to VLSI design to minimizing soap films to winning at Super Mario Bros. Meanwhile, another cluster of tens of thousands of problems has been proved to lie in P ... Those range from primality to matching to linear and semidefinite programming to edit distance to polynomial factoring to hundreds of approximation tasks ... To prove $P = NP$, it would suffice to find ... a single polynomial-time equivalence [between any of the NP-complete problems and any of the P problems] ... In half a century, this hasn't happened.

At the beginning of this section, I described **Conjecture Heuristic** as carrying the least philosophical baggage. I hope this seems reasonable now, considering what I have sketched is something all too familiar for the working mathematician: once a pattern occurs numerous times, from numerous sources, one is justified in conjecturing that this pattern will continue. As the Aaronson quote makes it clear, the usual tropes of prediction, inductive evidence, and empirical confirmation and fallibility[17] are all at work here, but at this point I should refrain from saying more, lest we be led astray from the main focus of the chapter.

Now, confluence was at least partially responsible for Church's and Post's conjectures

[16]Notice, however, that Kleene's argument by confluence appeals to positive evidence: all definitions have been proven to be equivalent; whereas here Aarsonson is appealing to negative evidence: none of the putative ramifications of $P = NP$ has ever been observed. It seems that appealing to a confluence of negative results is common in computational complexity theory; similar reasoning regarding the relativization barrier of P vs. NP can be found in Hartmanis et al. (1993) (attributed to Hopcroft): "No problem that has been relativized in two conflicting ways [e.g., for some oracles A, B we have $P^A = NP^A$ and $P^B \neq NP^B$] has yet been solved, and this fact is generally taken as evidence that the solutions of such problems are beyond the current state of mathematics."

[17]Section 4.2 of Copeland (2024) surveys skepticisms about these inductive and equivalence arguments in the case of the Church-Turing Thesis.

that certain definitions of effective calculability would be equivalent. Indeed, the repeated confirmation of such conjectures formed a firm basis of our acceptance of the later Church-Turing Thesis. Going one step further, common interpretation of the thesis entails that it does not really matter which definition one chooses. I will proceed to analyze this particular sentiment in the next two sections.

2.3 Rigor Assurance

The preceding section outlines the role that confluence plays in providing heuristic for conjectures. Mathematicians are well-accustomed with this general process, but admittedly it is not what immediately comes to mind when one thinks of Church's Thesis or the Church-Turing Thesis in actual practice. The nuts and bolts of proofs and theorems will be the focus of the present section.

Recall that Kleene's argument by confluence was meant to show that the analyses of Church, Turing, Post, and others provide the correct notion of effective calculability. Although confluence is considered strong evidence of this correctness, in practice the Church-Turing Thesis is typically invoked in a curious way, one that will reveal the role of confluence in this context.

As anyone familiar with computability theory will know, the Church-Turing Thesis is most frequently cited in the phrase "proof by Church(-Turing)'s Thesis". And the simple observation here is just that, in many instances, such appeals are not intended to say that we have uncovered new truths of the One True Notion of Computable. They are instead intended as warrants for something else. To see an example, consider the following passage from the first chapter of André Nies's influential textbook *Computability and Randomness*. Having defined computable functions in terms of Turing machines, Nies is quick to note that

> many other formal definitions for the intuitive notion of a computable function were proposed. All turned out to be equivalent. This lends evidence to the Church-Turing thesis[18] which states that any intuitively computable function is computable in the sense of [Turing machines]. More generally, each informally given algorithmic procedure can be implemented by a Turing program. We freely use this thesis in our proofs: we give a procedure informally and then take it for granted that a Turing program implementing it exists. (Nies, 2012, p. 3)

Clearly, Nies is making a characteristic appeal to confluence here. However, despite citing confluence as evidence for the Church-Turing Thesis, he is not, at least not in the present context, applying the thesis to promise the reader that the computability notion studied in the book is the One True Notion (Nies, 2012 is of course more of a mathematics textbook than a philosophy treatise). Rather, it is to anticipate a popular practice in the computability theory literature: that when it comes to proofs and arguments, informal language will be preferred over formal manipulations. Furthermore, the reader is assured that no loss of rigor will be incurred in doing so.

This is what I call **Rigor Assurance**, and a sizeable portion of the confidence granted by confluence (in the Church-Turing Thesis) is actually confidence for that. The following passage from Rogers is almost tailor-made to make this point:

> These techniques [of relying on informal descriptions of algorithms] have been developed to a point where (a) a mathematician can recognize whether or not an alleged informal algorithm provides a partial recursive function ... and where (b) a logician can go from an informal definition for an algorithm to a formal definition ... [S]uch methods ... permit us to avoid cumbersome detail and to isolate crucial mathematical ideas from a background of routine manipulation. We shall

[18]It is worth mentioning that this is the first place in the book where the term "the Church-Turing thesis" occurs.

> see that much profound mathematical substance can be discussed, proved, and communicated in this way ... Of course, [anyone arguing this way] ... must be prepared to supply formal details if challenged. Proofs which rely on informal methods have, in their favor, all the evidence accumulated in favor of Church's Thesis. (Rogers, 1987, p. 20)

Recalling Rogers's own expression of **Conjecture Heuristic** in the previous section, and considering that these two passages appear only two pages apart in Rogers's textbook, it is likely that (for Rogers, at least) the pull of **Rigor Assurance** is partially borrowed from the former: given the vast amount of equivalent definitions of "computable", one can be confident that informal descriptions of an algorithm can be formalizable as the formal algorithms of, say, Turing machines; and in the event that one's formalization departs in minute details from the usual definition of Turing computation, one is justified in believing that the difference is inconsequential, as long as the end result seems to obey reasonable criteria of what counts as computation[19].

The practice of proving something by Church's Thesis is well-documented in the literature. Its practical and philosophical merits are extensively studied, for example, in San Mauro (2018), which will serve as the occasional interlocutor in the remainder of this section and the next. My goal in doing so is to leverage certain subtleties in San Mauro's analysis, which will help me draw out and clarify the distinction between what I call **Rigor Assurance** here and **Coding Invariance** later, a finer distinction that is often overlooked in the literature.

To start, San Mauro conspicuously calls the aforementioned practice "the practical side of Church-Turing Thesis", referring namely to the collection of all the appeals to the thesis that appear in some steps of some mathematical proof[20]. Although I will eventually argue

[19] In the present context, this serves only to warrant trust in using informal language in proofs. But I shall claim that this invariance under different choice of formalisms is itself an additional facet to confluence, which will be the theme of the next section.

[20] I refer the reader to standard computability theory textbooks or to San Mauro (2018) for examples.

that there is an additional dimension to this seemingly uniform practice, let me cite San Mauro's identification of the following principle allowing rigor to be assured:

(CTB) If any informal description of an algorithm can be formally implemented in each model of computation (e.g., assuming the Church-Turing Thesis) then, in order to prove that something is computable, it is sufficient to describe an informal way to compute it and then make reference to the Church-Turing Thesis.

For an instance of (CTB) in practice[21], here is Goldbring and Hart forecasting their choice of the informal:

> There are many approaches to formalizing [computability] (e.g. Turing-machine computable functions and recursive functions) and all known formalizations can be proven to yield the same class of functions. This latter fact gives credence to the Church-Turing thesis, which states that this aforementioned class of functions is indeed the class of functions that are computable in the naïve sense described above. In the rest of this paper, we will never argue about this class of functions using any formal definition but will only argue informally in terms of some kind of algorithm or computer; this is often referred to as arguing using the Church-Turing thesis. (Goldbring & Hart, 2021)

Presumably, Goldbring and Hart (2021) is not primarily concerned with discovering what is and what is not computable in the intuitive sense, the way perhaps the earliest works on the Entscheidungsproblem or Hilbert's tenth problem were. It merely treats computability as a tool to solve problems in the model theory of operator algebras. In other words, the role played by the Church-Turing Thesis in their paper is merely to assure rigor and simplify

[21] Supplementing San Mauro's comment that "the most immediate source of observations regarding how such practice has to be intended comes from the kind of expository remarks [in textbooks]", we will see plenty of such instances in the specialist journal articles too.

presentation. The evidence from confluence accrued for the Church-Turing Thesis is being used to support the belief that this type of informality will not lead to error.

And indeed, to assure rigor, there is no need to appeal to the Church-Turing Thesis at all. Consider historically the first paper relating computability and measure theory. The article Leeuw et al. (1956) is concerned with machines equipped with access to outcomes of random experiments and aims to show how a machine may or may not be enhanced to take advantage of them[22]. Before delving into the technical arguments, the authors announce:

> [The proofs in the paper] can be considered to be indications of how the theorems, if they were stated formally in the language of recursive function theory, could be proved formally using the tools of that theory. This formalization is not carried out in the paper since it would detract from the conceptual content of the proofs. However, it should be clear to anyone familiar with recursive function theory that the formalization can be carried out.

Let us focus now on de Leeuw et al.'s justification of using informal natural language in their proofs. Here we see the same kind of **Rigor Assurance** as in Nies's textbook and Goldbring and Hart's article, this time without explicitly invoking the Church-Turing Thesis. To the authors, formalization serves more to distract than to illuminate, and hence it ought to be kept at minimum. Furthermore, doing so is safe, because the vast body of work in computability theory has assured the reader that they can trust that such formalizations are possible.

For Hamkins and Lewis, the ability to translate from the informal to the formal is a transferable skill, to context more detached from ordinary computability. In pioneering works of infinitary computation, **Rigor Assurance** allowed the authors to take the high road:

[22] We will return to these technical results in a later section.

> We will assume complete familiarity with the notions of Turing machines and ordinals and, in describing our algorithms, take the high road to avoid getting bogged down in Turing machine minutiae. We hope the readers will appreciate our saving them from reading what would otherwise resemble computer code. (Hamkins & Lewis, 2000)

In Carl's survey of the later works of infinite-time decidable equivalence relations, this is considered a virtue: "...ITTM-reducibility has the advantage [of having] rather intuitive informal descriptions which can be used in proofs ... One may say that 'proof by Church-Turing thesis' is more readily available for ITTM-reducibility than for Borel reducibility." (Carl, 2019, p. 263)

To really highlight the role of confluence in **Rigor Assurance** (and to see that this role can really be decoupled from other commitments brought by the Church-Turing Thesis), consider α-recursion theory. Greenberg (2020) opens with a survey of the array of equivalent definitions of admissible computability[23] and an intricate network of related facts showing how α-computability shares many of the analogous results of ordinary computability. Having primed readers of the multi-faceted nature of α-computaility, Greenberg writes:

> In general, working in α-computability, with experience, we apply some kind of Church-Turing thesis to α-computable functions ... we eventually cease to write down precise Σ_1 formulas ... Instead, we develop an intuition as to what constitutes "legal" α-computable manipulations of α-finite objects (elements of L_α), and get a sense of the "time" that a process takes; if it takes fewer than α steps, then it "halts".

[23]For example, for an admissible ordinal α and $A \subseteq \alpha$ the following are equivalent: 1. A is Σ_1-definable in L_α. 2. A is computably enumerable by Koepke's α-Turing machines. 3. A is semi-decidable by Koepke's α-register machines. See Koepke and Seyfferth (2009) and Carl (2019, Theorem 3.3.3) for definitions and proofs.

What “kind of Church-Turing thesis” is at work here? The “intuition” that Greenberg says the reader develops is certainly not one about some far-fetched notion of computability that deals with the transfinite, waiting to be captured by α-computability, but rather, it is one about how to translate informal language used in proofs to formal arguments. Perhaps more accurately, having seen the confluence of definitions for α-computability and the appropriate generalizations of the familiar results from ordinary computability, the reader is expected to develop confidence in the use of informal language. In the absence of a conspicuous informal notion to be captured by α-computability, Greenberg’s appeal to the “kind of Church-Turing thesis” in studying the α-computable functions can only be interpreted to mean an appeal to **Rigor Assurance**.

These examples attest to the ubiquity of **Rigor Assurance** in practice. It is perhaps the most practical invocations of the Church-Turing Thesis in the technical literature. I hope to have shown that many of these are not really invocations of the received wisdom of the Church-Turing Thesis, but means of crossing San Mauro’s Church-Turing Bridge.

Much of the remainder of San Mauro (2018) is concerned with adjudicating a controversy about the significance of this practice. San Mauro identified a tension between the ubiquity and usefulness of this practical side of the Church-Turing Thesis on the one hand, and on the other hand the question of whether this practical side is any more than just a matter of routine convenience.

San Mauro calls the following the Standard View (SV) and attributes it to his interlocutors (San Mauro, 2018, p. 236):

SV(a). The Church-Turing Thesis allows us to rely on informal methods (by (CTB));

SV(b). Yet, these methods are in the end just a matter of convenience: informal definitions point towards formal ones, and we could theoretically substitute the former with the latter without any significant loss or gain of information;

SV(c). This operation is analogous to what happens in most parts of mathematics.

San Mauro's interlocutors are authors who do not see much difference between "proof by Church's Thesis" and other routine mathematical drudgeries. Odiffredi, for example, calls **Rigor Assurance** "avoidable ... does not even require a Thesis: it is just an expression of a general preference, widespread in mathematics, for informal (more intelligible) arguments, whenever their formalization appears to be straightforward, and not particularly informative." (Odifreddi, 1989; as cited in San Mauro, 2018)

He also describes Epstein and Carnielli as the proponents of (SV), who have judged in their computability textbook that "To invoke Church's thesis when "the proof is left to the reader" is meant amounts to giving a fancy name to a routine piece of mathematics while at the same time denigrating the actual mathematics." (Epstein & Carnielli, 2008; as cited in San Mauro, 2018)

I should pause here and point out that San Mauro is not being entirely fair to Epstein and Carnielli[24]. Here Epstein and Carnielli is responding to Daniel Cohen, who described Church's Thesis as "... more than a philosophical statement about the nature of computability. It is a useful tool in proofs. To show that the function is partial recursive [formally] ... [is] not likely to give any insight ... [and is] usually replaced by an appeal to Church's Thesis." (Cohen, 1987, p. 104; as cited in Epstein & Carnielli, 2008, p. 231). In my view, San Mauro should have attributed the Standard View (SV) to Cohen, who argues most forcefully for it in the quoted passage, and not to Epstein and Carnielli, who are merely pointing out that it is mistaken to call this practice "invoking Church's Thesis", especially when the latter is in fact much more involved. At any rate, both Odiffredi's and Cohen's sentiments are concrete demonstrations of the central thesis of the present chapter, that many invocations of the Church-Turing Thesis are really appeals to something else much less Thesis-looking, and

[24]Epstein and Carnielli enthusiastically call the equivalences of definitions of computability *The Most Amazing Fact* in Epstein and Carnielli (2008), which commendably dedicates a large portion to the discussion the Church-Turing Thesis and its attendant evidence and philosophies of mathematics.

Epstein and Carnielli is just one step short of making this point: Cohen is indeed pointing out the salient practice of invoking Church's Thesis in proofs, but what he is referring to is merely **Rigor Assurance** in disguise.

San Mauro (2018) eventually arrives at the diagnosis that (SV) omits a key part of computability and hence should be rejected: "SV fails to represent a central phenomenon of Computability, that of conceiving most constructions as absolute, i.e. independent from the background formalism and yet not to be regarded as incomplete." (San Mauro, 2018, p. 246) To make this point, San Mauro considers the proof of the existence of a simple set[25]. As with most proofs in computability theory, the proof begins with a numbering of the partial computable functions (i.e., a c.e. surjection from natural numbers to programs), and then proceeds to construct a simple set from this numbering. San Mauro notes that the proof does not specify the background enumeration on which the proof is based, and yet the proof is considered complete, because a set remains simple under all acceptable (in a technical sense) numberings.

And this places considerable burden on proponents of (SV), who would claim that, barring further details like what coding is used, such a semi-formal proof "has to be intended as a sort of a prototype, to be completed by specifying a certain numbering" and "would correspond to a general method for describing, for each acceptable numbering, a corresponding simple set". This view is mistaken, says San Mauro:

> ... not specifying the background numbering ... subsumes the kind of practical use of [the Church-Turing Thesis] that we have extensively discussed ... The theory of simple sets is invariant with respect to the acceptable numbering we choose to work with, making this very choice superfluous. So, against (SV),

[25] A simple set is a co-infinte, computably enumerable set of natural numbers that meets every infinite, computably enumerable set. The reader is not required to know the definition or the proof, only that the existence of a simple set is not a trivial fact, and that simplicity is invariant under different choice of enumerating the computably enumerable sets.

> ... [the simple set constructed] does not refer to any of its formal definitions ... [and] although our informal proof does provide a method for producing, for all numberings, a given simple set ... [T]o collapse the meaning of such proof to this method would correspond with claiming that an implicit reference to numberings is somehow needed to make complete sense of our construction ... Rather, the notion of simplicity is better understood as an absolute one, i.e. independent from the chosen formalism. (ibid., pp. 242-243)

I see the tension identified by San Mauro as a disagreement about commitment. For San Mauro, invoking Church's Thesis in a proof commits its participants (i.e., its writers and readers) to more. His point is that, by invoking the Church-Turing Thesis in a proof, we see ourselves as dealing with an absolute notion of computability (The One True Notion), instead of merely backing away from the cumbersome details of formalization and leaving them to the reader. Most investigations in computability theory are like this (that is, studying objects invariant under choice of coding). Rogers (1987), for instance, seems to agree that Church's Thesis is committing us to dealing with The One True Notion, when he wrote the following after introducing the concept of Gödel numbering of programs:

> The use of codings raises an immediate question of invariance. Once a coding is chosen, will the formal concept partial recursive function on code numbers correspond to the informal notion algorithmic mapping on the uncoded expressions? As the latter notion is informal, the answer must be, in part, empirical. Church's Thesis provides an affirmative answer. (p. 28)

Proponents of (SV), on the other hand, are characterized as thinking there is no such commitment, seeing as they reduce the practice of invoking Church's Thesis to a matter of routine convenience with no additional significance. And in doing so, they miss out on this key formalism-free aspect of computability.

Contrary to San Mauro's characterization, however, I claim that "not specifying the background numbering" should not subsume "the kind of practical use of the Church-Turing Thesis" considered so far (i.e., **Rigor Assurance**), and that the quoted passage from Rogers above is in fact pointing to a different facet of confluence, a kind of **Coding Invariance**. Furthermore, I shall argue that the tension between the practical side of the Church-Turing Thesis and proponents of (SV) arises from a misalignment of what role each party is expecting Church(-Turing)'s Thesis to play; or better yet, what each party is expecting confluence to justify. Both sides are right, and there is in fact no serious disagreement. I will elaborate this point in the next section.

2.4 Coding Invariance

By way of transition, let us notice that Rogers is making a subtly different point in the quoted passage above. Recall when he promised **Rigor Assurance**, he meant "a logician can go from an informal definition for an algorithm to a formal definition ... [S]uch methods ... permit us to avoid cumbersome detail." This is assurance that some formalization exists and is easily available, whearas his remark on codings guarantees that it does not matter which one.

Kleene (1952) seems to assume that these are two different facets of confluence. Having made his argument by confluence, he lists the stability under different codings as a separate piece of evidence for Church's Thesis: "Of less weight, but deserving mention, is the circumstance that several formulations of the main notions are equivalent; i.e. the notions possess a sort of 'stability'." (p. 320)

Kleene proceeds to cite the proofs of equivalences of different ways of formalizing the same idea. I should stress that, crucially, Kleene is not citing (say) the equivalence between general recursiveness and λ-definability here - he has already made that argument before - he is citing the equivalences of (say) different variants of λ-definability:

> The notion of λ-definability has the variants λ-K-definability ... and λ-δ-definability ... also there is a parallel development, started by [Schönfinkel, Curry, and Rosser], which leads to a notion that we may call combinatory definability, proved equivalent to λ-definability by Rosser. (p. 321)

Kleene would go on to mention similar developments in other families of formalizations. But his main point is clear: invariance[26] under the choice of codings is different from guarantee that formalization is possible in principle.

Why does this invariance matter? In classical computability theory, just as in its higher cousin hyperarithmetic theory, a common practice is to talk about things in terms of their codes. This is in some sense inevitable, because things like ordinals or well-orderings are not a native part of the theory's vocabulary. So one finds need to utilize codings for constructive ordinals or the H_a in hyperarithmetic theory[27]. And for Moschovakis (2016b), such uses provide additional insight:

> Codings are useful for expressing succinctly uniform properties of coded sets. Their general theory is technically messy, not very interesting mathematically and certainly not worth putting here ... It is clear that propositions like Lemmas 5.3.1 and 5.3.2 which hold uniformly for a certain coding also hold uniformly for every equivalent coding — and for some of them the proof might be easier[17] [Moschovakis's footnote]. We exploit this idea by establishing an elegant characterization of [the class of hyperarithmetic sets] which produces a coding for it

[26] Or stability in Kleene's term. The word "stability" is very loaded, as we shall see later, which is why I chose not to adopt it.

[27] See Moschovakis (2016b) for details. The reader only need to know that these are fundamental concepts in hyperarithmetic theory.

> equivalent to the classical one in (5.23) but much simpler. (pp. 120-121)

The latter part of Moschovakis's remark, which he conspicuously titled **Coding Invariance**, is further explained in footnote 17, where he references exactly the kind of "central phenomenon of Computability" that San Mauro champions:

> For a classical example, consider the coding of recursive partial functions specified by [Kleene's Normal Form Theorem]. Its precise definition depends on the choice of computation model that we use, Turing machines, systems of recursive equations or whatever [that is, each choice of model gives rise to a different enumeration of the recursive partial functions], but all these codings are equivalent and so uniform propositions about them are coding invariant. (ibid.)

Note also, that by "technically messy, not very interesting mathematically", Moschovakis is certainly not referring to the kind of informal-to-formal translation that proponents of (SV) are claiming to be messy and routine. A brief glance at Moschovakis (2016b) will show that he makes very little use of informal language characteristic of classical computability theory, mostly because he has no need to describe any algorithm at all! So the worry here is certainly not about the validity of proofs (something that would otherwise be assuaged by **Rigor Assurance**), but that we might be led to pseudo-insights that are really just byproducts of the peculiarities in our choice of coding.

Immediately following Moschovakis (2016b) in the same collection is Boker and Dershowitz (2016), who take the perils of coding peculiarities as a point of departure and proposed a framework for characterizing them. Boker and Dershowitz motivate this issue of *honest* versus *dishonest* coding with a few examples. To name one, consider the usual practice in (theoretical) computer science of encoding a graph as some kind of data structure, e.g., a list of nodes and edges. The point Boker and Derschowitz make is that if one insists or happens to encode a graph by listing the nodes in the order of a Hamiltonian path (if one exists), then deciding the existence of such a path, a classic NP-complete problem, can be solved in linear time. This is a dishonest coding, because it is the choice of encoding that makes the problem easy. And in fact, this will have the unwelcomed consequence that one "will only be able to quickly answer the specific question whether there is a Hamiltonian path, whereas she would have a much harder time performing basic graph operations, such as adding an edge."

They go on to make a point strongly resembling Moschovakis's, that ultimately **Coding Invariance** safeguards us from such unncessary worries:

> The [complexity] of a problem turns out to be essentially independent of the particular encoding scheme and computer model used for determining time complexity ... In fact, the standard encoding schemes used in practice for any particular problem always seem to differ at most polynomially from one another. It would be difficult to imagine a "reasonable" encoding scheme for a problem that differs more than polynomially from the standard ones. (Garey & Johnson, 2009; as cited in Boker & Dershowitz, 2016, p. 153)

Appealing to this **Coding Invariance** aspect of confluence is exactly how Sipser dispels worries about choice of computation models after introducing the complexity class P in his classic textbook on computational complexity:

> In complexity theory, we classify computational problems according to their time complexity. But with which model do we measure time? The same language may have different time requirements on different models.
>
> Fortunately, time requirements don't differ greatly for typical deterministic models. So, if our classification system isn't very sensitive to relatively small differ-

> ences in complexity, the choice of deterministic model isn't crucial ... All reasonable deterministic computational models are polynomially equivalent. That is, any one of them can simulate another with only a polynomial increase in running time. (Sipser, 2013, p. 286)

A moment of reflection is needed here. The existence of universal enumerating programs is not due to some whimsical definition of a program or a Turing machine, which might smuggle in a particular way of encoding, say, the computable functions. It is a substantive (or absolute, in San Mauro's terms) property of the partial computable functions that there is one that enumerates all of them. While this is most readily seen by appealing to Turing machines, it does not technically depend on identifying these functions as the ones implemented to Turing machines specifically; any model of computation will allow us to prove the same thing. To summarize the difference in somewhat slogan form: **Rigor Assurance** guarantees no rigor is lost in translation. **Coding Invariance** guarantees no extraneous peculiarities is added in codification.

This key insight allows us to resolve the tension identified by San Mauro (2018). Recall that the proponents of (SV) claims that "proofs by Church's Thesis" is a completely routine and humdrum practice, nothing more significant than saying "the proof is left to the reader." San Mauro, on the other hand, argues that (SV) fails to capture the central phenomenon of computability, that of considering most constructions as absolute, i.e., invariant under the choice of codings. I claim this is merely a misunderstanding: **Rigor Assurance** with the use of informal language is of course routine and unremarkable, it is simply the kind of thing that licenses professional mathematicians to choose not to spell out every last detail of a proof or leave it to the reader. **Coding Invariance**, on the other hand, is a much more profound and substantive property of computability, one that is characteristic of it and not many other fields of mathematics (at least not in every one where someone has left an exercise to the reader). And it is this that is being invoked when we say that the proof of the existence of a simple set is complete without specifying the background numbering.

Thus, the tension identified by San Mauro stems from a misalignment of expectations on both parties. They misalign on what role they take Church(-Turing)'s Thesis to play. Proponents of (SV) are right to consider routine and unremarkable the kind of **Rigor Assurance** obtained by invoking Church's Thesis, but they are too quick to think that this is all there is to it. San Mauro is right to see the practice as committing us to the substantive **Coding Invariance** characteristic of computability, something that is significant, but that this is not what proponents of (SV) are denying. As we have seen, there is no serious disagreement between the two parties, and the tension is resolved by recognizing that the Church-Turing Thesis is being invoked for two different reasons; or better yet, confluence is being appealed for two different justificatory purposes.

Returning to the matters at hand, we observe that **Coding Invariance** extends beyond just the computable and the hyperarithmetic. Recall the following Spector-Gandy-type characterization of Σ^1_2 sets, with which Chong et al. (2019) set up the rest of their paper:

Theorem 2.1 (Theorem 1.1 in Chong et al. (2019))**.** *Given a set $A \subseteq 2^\omega$, the following are equivalent*

1. *A is Σ^1_2, that is, definable in second-order arithmetic by a formula of the form $\exists f \forall g \varphi$, where f, g are real number variables and φ only quantifies over the naturals.*
2. *There is a Σ_1-formula φ such that for all reals x, $x \in A \iff L_{\omega_1^{L[x]}}[x] \models \varphi(x)$.*
3. *There is a Σ_1-formula φ such that for all reals x, $x \in A \iff L_{\delta^1_2(x)}[x] \models \varphi(x)$, where $\delta^1_2(x)$ is sup of the ordertypes of $\Delta^1_2(x)$ well orderings of ω.*

Having reminded the reader of the classic theorem that will be used throughout their paper, Chong et al. make an appeal to a Church-Turing-type thesis: "Theorem 1.1 enables one to use recursion-theoretic arguments to study Σ^1_2-sets. It follows that there is a version of 'Church-Turing Thesis' for Σ^1_2-sets that we can appeal to in the construction of such sets."

Chong et al. (2019) serves as yet another instructive example in distinguishing **Rigor Assurance** and **Coding Invariance**. Here, again, Chong et al. are not anticipating the use of informal description of algorithms. They are instead preparing the reader for their liberal context-switching, utilizing different codings or representations of the Σ^1_2 sets as they see fit. Granted, this confluence at Σ^1_2 licenses the authors to talk freely about, for example, "effectively computing an index of a Π^1_1 set" and "fix[ing] a L_{ω_1}-effective enumeration of perfect trees". This much is **Rigor Assurance**, similar to Greenberg's application of his "Church-Turing Thesis" for admissible computability (2020). But the key point here is not that the reader needs assurance that such use of informal language will not render the proofs invalid. Quite on the contrary, the reader is promised that translation is always available, not between informal language and formal language, but between "languages" from different fields of mathematics, so the proof is not led off-topic and end up talking about something else entirely.

For another example of how coding peculiarities might smuggle in pseudo-insight, consider the following well-known fact in proof theory and recursion theory[28]. The theorem concerns transfinite progressions of consistency statements. Noting that one can only talk about limit stages of such progressions by talking about codes of computable well-orderings, we define $T_0 := \mathsf{ZFC},\ T_{2^n} := T_n + \mathrm{Con}(T_n),\ T_{3\cdot 5^e} = \mathsf{ZFC} \cup \bigcup T_{\Phi_e(n)}$. Now coding peculiarity follows:

Theorem 2.2 (Turing's completeness theorem[29])**.** *For every true Π^0_1 sentence φ, there exists a notation in d in Kleene's $\mathcal{O}$ such that φ is provable in T_d.*

Roughly, if we want to study the strength of iterated consistency statements, then the strength of the limit theories are extremely sensitive to the choice of codings. Every true Π^0_1 sentence is provable in some T_d, a sensible candidate for encoding a limit theory. This is an

[28] The standard reference is Feferman (1962). See also the recent preprint Pakhomov et al. (2024)

[29] Again, unrelated to the other notion known as Turing completeness.

obvious lack of **Coding Invariance**. The same coding peculiarities also underlie the famous Kreisel's Counterexample (cf. Theorem 7.5.1 in Pohlers (1989)) attesting to the delicacy in treating the notion of proof-theoretic ordinals.

This facet of **Coding Invariance** is at play in a particularly amusing episode in the history of quantum computation. In one of the earliest papers defining quantum Turing machines and its complexity classes, appearing in *Proceedings of the twenty-fifth annual ACM symposium on Theory of Computing* (Bernstein & Vazirani, 1993), the authors made no restriction on what transition amplitudes were allowed. This enabled Adleman et al. (1997) to prove that "the class of sets which are decidable [by Bernstein and Vazirani (1993)'s machines] with bounded error in polynomial time has uncountable cardinality and contains sets of all Turing degrees." This is certainly an artifact of coding peculiarity, and the culprit was quickly identified and corrected - in the journal version (Bernstein & Vazirani, 1997) of Bernstein and Vazirani (1993) added a uniform computability requirement to what quantities are allowed as transition amplitudes, acknowledging Adleman et al. (1997), which was published in the same journal:

> we need to constrain the entries allowed in the transition function of our probabilistic TM. Otherwise, it is possible to smuggle hard-to-compute quantities into the transition amplitudes, for instance by letting the ith bit indicate whether the ith deterministic TM halts on a blank tape ... As in the case of probabilistic TM, we must limit the transition amplitudes [of quantum Turing machines] to efficiently computable numbers. (Bernstein & Vazirani, 1997, p. 1417)

Careful choices were then made to ensure that, in terms of computability (i.e., what are computable simpliciter), the resulting machines are equivalent to classical Turing machines, citing Adleman et al. (1997) and an unpublished manuscript by Solovay and Yao. The resulting definition of quantum Turing machines is now the canon. This prompts Peter

Shor[30] to ask if a mathematical definition can be wrong (Shor, 2020). A cursory look at what respondents had to say in that question reveals many a faint echo of what I call **Joint-carving**, **Coding Invariance**, and **Remarkable Coincidence**.

At any rate, I hope to have convinced the reader that **Coding Invariance** is a key facet of justification by confluence, and sufficiently distinct from **Rigor Assurance**. Despite not making the distinction, San Mauro (correctly, in my view) elucidates this facet of **Coding Invariance** by citing Burgess's notion of *indifferentism* (J. P. Burgess, 2008, 2015; see also Kennedy, 2020, for extended discussions and further examples). For Burgess,

> the general phenomenon of the indifference of working mathematicians to certain kinds of decisions that have to be made in any codification of mathematics ... two analysts who wish to collaborate do not need to check whether they were taught the same definition of 'real number'. (J. P. Burgess, 2008, 2015; as cited in Kennedy, 2020, p. 6)

In the theory of Borel equivalence relations, Gao's thesis is an epitome of confluence-driven indifferentism. The rich theory of Borel equivalence relations aims provide a complexity hierarchy to the various classification problems that arise in other areas in mathematics. The theory, also known as invariant descriptive set theory[31], incorporates tools from effective and classical descriptive set theory, and, making substantial use of measure and category machinery, applies them to show how one classification problem is reducible or irreducible to another.

In the literal sense, what is being proved is the (non-)existence of Borel (or otherwise definable) reductions[32] between Borel equivalence relations on Polish spaces. The significance

[30] whom I thank for this curious bit of history and the exact references.

[31] The eponymous textbook Gao (2008) covers the state of the art, to which we refer the interested reader. We will presently be concerned with what Gao has to say about the practice of invariant descriptive set theorists.

[32] Let E be a Borel equivalence relation on a standard Borel space X and similarly F a Borel equivalence

of such technical results crucially depends on the choice of coding: the choice of topology, which provides a Borel structure, and the choice of how specific mathematical objects are coded as elements of certain Polish spaces. In the absence of any sort of motivating story like Turing's, there is no prior safeguard against coding peculiarities.

In practice, however, this choice rarely matters. Reaching the end of the book and reflecting on the vast amount of application to concrete classification problems, Gao summarizes, having just finished a long proof of "the satisfactory correspondence" between two such actual choices:

> The existence of such an isomorphism means that the two approaches to equip [the space in question] with a standard Borel structure are equivalent. Note that it is mathematically nontrivial to prove this equivalence. However, in practice, whenever we have different approaches to equip standard Borel structures on hyperspaces ... they always end up to be equivalent despite the difficulty of the proof. (Gao, 2008, p. 328)

This motivates Gao's Thesis[33]:

(Gao's Thesis). For any class $\mathcal{H}$ of mathematical structures, if (X_1, Ω_1) and (X_2, Ω_2) are two standard Borel spaces naturally coding elements of $\mathcal{H}$, then there exists a Borel map $f : X_1 \to X_2$ such that $f(x)$ and x are isomorphic as mathematical structures for every $x \in \mathcal{H}$.

For a light application of Gao's Thesis, consider Gao's indifference to what he takes

relation on a standard Borel space Y. We say E is *Borel reducible* to F (written as $E \leq_B F$) iff there is a Borel function $f : X \to Y$ such that $x_1 E x_2 \Leftrightarrow f(x_1) F f(x_2)$. Such a function f is called a *Borel reduction* of E to F.

[33]The name Gao's Thesis is given by Hamkins (Hamkins, 2016). The version here is taken from Kaya (2016, p. 35), a version that both Hamkins and Kaya agree stays closer to how the thesis is actually used in practice.

countable groups to be when he studies reductions among equivalence relations between them: "A majority of our equivalence relations are defined on $2^{\mathbb{N}}$ or some variations such as $2^{\mathbb{N}\times\mathbb{N}}$. We view elements of this space to be either functions from $\mathbb{N}$ to $2 = \{0, 1\}$ or subsets of $\mathbb{N}$;and we freely switch our point of view without explicit mention." (Gao, 2004). Many other such examples can be found in Kaya (2016, Chapter 3).

In sum, Gao expresses a sentiment about **Coding Invariance** which we should now be thoroughly familiar with:

> For readers familiar with computability theory, the statement is similar to the Church-Turing Thesis ... To a large extent such work [i.e., proving equivalences between coding] is mathematically insignificant and irrelevant to the understanding of the main problem. For our purpose the objective is to understand the complexity of various classification problems for, say, Polish metric spaces. While the theorems in this section are nontrivial and their proofs contain important ideas that can be used later to tackle the classification problems, the statements of the theorems themselves are not helping in our understanding of the complexity of the classification problems. (Gao, 2008, p. 328)

By now we have seen the same attitude already expressed by Moschovakis, Garey and Johnson, and to some extent Boker and Dershowitz. Contrast this with Gao's patent appeal to **Rigor Assurance** in the same textbook, earlier on when covering the computability preliminaries, anticipating the use of informal language:

> All formal definitions of computability have been shown to be equivalent to the above one [i.e., general recursiveness]. In fact, when working with computable functions we will not deal with the details of the above definition, but will rather adopt the Church-Turing Thesis, which states that any of the formal definitions

> of computable functions captures the intuitive notion of computability. Thus if a function is intuitively computable by an informal algorithm then by the Church-Turing Thesis we may conclude that it is formally computable without checking the details of the formal definitions. (Gao, 2008, p. 24)

I take this contrast to be decisive evidence that **Rigor Assurance** and **Coding Invariance** are definitively distinct, although they are sometimes conflated in the case of classical computability. Once the distinction is recognized, the tension identified by San Mauro can, as I have argued above, be satisfactorily resolved by seeing that both parties are right in their respective understanding of what confluence is and is not doing.

Now, the importance of Gao's Thesis motivated Hamkins (2016) to ask for a counterexample on the one hand (this remains unanswered), and for a philosophical analysis for how Gao's Thesis is analogous to the Church-Turing Thesis, as well as whether one can provide a Turing-esque philosophical grounds in support of the thesis. We have partially addressed this latter question. The next chapter will provide more conceptual tools to figure out how the two theses are alike exactly, and how they are not. Eventually I shall argue that such a Turing-esque motivating story is not necessary, and the empirical observation of these equivalences alone suffices for the purpose at hand, as Gao's Thesis does not engender all the same kinds of philosophical commitments as the Church-Turing Thesis, but only this very practical aspect of **Coding Invariance**.

I have to tried to characterize **Conjecture Heuristic**, **Rigor Assurance**, and **Coding Invariance** as facets of justification by confluence that are relatively light on philosophical commitments. However, in addition to being a versatile practical tool and a safeguard for formal rigor, confluence can also be very enticing philosophically. For example, the Tarski-Sher Thesis maintains that logicality is to be identified with invariance under all isomorphisms of the relevant domain (Sher, 2008; see also Chapter 5.4 of Kennedy, 2020, for an extensive overview of critiques and supports). Many authors have put the words "fundamental", "robust", "stable", and "thesis" within in same paragraph where they reference some sort of confluence. The next chapter will be about these more philosophically loaded aspects of confluence.

Chapter 3

Justification by Confluence, Part II: Philosophical Matters

3.1 Joint-carving

Introductory classes in psychology routinely advise presenters to place at the initial and final positions those points they want remembered. Content more familiar to the audience ought to go in the middle, for familiarity will counterbalance the disadvantages of this medial placement. With apologies to the reader, this is the strategy I am following here.

The formalization of computability, as exemplified by the pioneering works of Turing, among others, marked one of the most prominent accomplishment of conceptual analysis in the last century. One reason for its success, as well as one of its distinguishing features, lies in the equivalence of the numerous formalisms independently devised by separate authors, with their respective motivations. Unsurprisingly, this lends support to its most significant philosophical import, recalling Gödel's remarks in his 1946 Princeton Bicentennial lecture "With this concept one has succeeded in giving an absolute definition of an interesting epistemological[1] notion, i.e., one not depending on the formalism chosen."

Implied in this remark is the idea that Turing had managed to give a *correct* formal definition of informal computability. Again, this is the received wisdom of the Church-Turing thesis. To borrow a classic phrasing, Turing's analysis, along with the confluence of definitions, managed to carve nature at its joint: it carves out the correct collection of functions/sets that correspond to the intuitively computable ones. This is an example of the **Joint-carving** role played by confluence. In the context of the Church-Turing thesis, it manifests in the common conclusion that Turing obtained the correct characterization of computability, bolstered by the compelling evidence that other independently motivated formalizations turn out to be equivalent. The fact that these distinct formalizations yield the same set of computable functions lends support to the idea that he has indeed captured the defining features of computation.

Turing's conceptual analysis of computation began from first principles: how an idealized human computor would compute, and from there, what they could and could not compute. Kleene, whose argument by confluence we have referenced a few times, includes this as a separate piece of evidence that Church's Thesis was correct. Kleene's point was that, among all the equivalent definitions of computability, there is a privileged one, to which all others were shown to be equivalent to it.:

> Turing's computable functions (1936-7) are those which can be computed by a machine of a kind which is designed, according to his analysis, to reproduce all the sorts of operations which a human computer could perform, working according to preassigned instructions. Turing's notion is thus the result of a direct attempt

[1] In Gödel (1986) the same notion (as well as two others, see below) was referred to as "metamathematical". Elsewhere, commenting on Skolem's first proof of a Löwenheim-Skolem-like theorem, he observed a "...lack of the required epistemological attitude toward metamathematics...", where the occurrence of "epistemological" was essentially omissible. Given Gödel's alternating adjectives for the three notions and his rather liberal use of the word "epistemological", it is unlikely that Gödel was using the word "epistemological" here in a serious way philosophically.

to formulate mathematically the notion of effective calculability. (Kleene, 1952, p. 321)

By now we have seen lots of examples of appeals to confluence. Many of them come from areas outside of classical computability. The privileged position of computability as a pinnacle of successful conceptual analysis is entailed in part by Turing's work. Indeed, Turing's conceptual analysis, the confluence of definitions, and the success of the Church-Turing Thesis have made a popular case study in the philosophy of science and mathematics, in relation to Carnapian explication and conceptual analysis (De Benedetto, 2021; Quinon, 2021; Rescorla, 2007; Sieg, 1997; and also Chapter 3 of Kennedy, 2020, just to name a few). Kennedy, for example, characterizes Turing's work as providing a *grounding*: "[besides the equivalences,] there is also grounding: the idea that, among the class of conceptually distinct precisifications of the given (intuitive) concept, one of them stands out as being indubitably adequate - as being the right idea." (Kennedy, 2020, p. 42)

This important feature of computability, i.e., the availability of a formalism-free conception via Turing machines (Kennedy, 2013), is arguably the most philosophically significant feature of computability. On the one hand, assuming there is a robust, informal, pre-theoretic notion to be explicated to begin with, the fact that several independently motivated definitions converge speaks to the correctness of such definitions. On the other hand, in the face of the many equivalences, a prevailing motivating story (and formalism-freenesss) of one such definition seems to explicate why confluence occurs: it is *no accident* that we hit upon the same definition again and again, because we have laid our fingers on something that is there. With Turing's treatment of the idealized human computor, one is convinced that it is this concept that Turing captures that the definitions are converging on.

The preceding sketch is awkwardly vague and naïve, and only supposed to be a toy example of a broader practice that we see in the literature. For our present context, we will focus less on this formalism-free aspect of Turing's grounding, but more on this feeling of *no-accident* that it elicits. Formalism-free or not, what Turing gave was a distinguished analysis that motivated his formal definition. Someone convinced by Turing's conceptual analysis will view his definition of computability as natural and correct, and an explanation for why all other proposed definitions converged. In fact, a surprising source of this pattern of **Joint-carving** by confluence can be found in a rather foreign landscape than what we have surveyed. Let us temporarily leave the comfort zone of pure mathematics and venture into the murky waters of empirical sciences.

Nozick's Presidential Address at the 94th Eastern APA meeting (Nozick, 1998) began with three "strands to our ordinary notion of an objective fact or objective truth," the first of which is the ability to bring about confluence: "An objective fact is accessible from different angles. It can be repeated by the same sense (sight, touch, etc.) at different times, it can be repeated by different senses of one observer, and by different observers. Different laboratories can replicate the phenomenon." He then went on the highlight the role of invariance in physics:

> That invariance is importantly connected to something's being an objective fact is suggested by the practice of physicists, who treat what is invariant under Lorentz transformations as more objective than what varies under these transformations. Dirac writes, "The important things in the world appear as the invariants... of ... transformations."

I believe part of what makes **Coding Invariance** an attractive feature of computability, and what motivated San Mauro and others to say that it reveals the absoluteness of computability notions, is this feeling that we have put our finger on something real, and real things tend to be invariant under different codings, to paraphrase Dirac[2]. Planck, for

[2]Epstein and Carnielli (2008) makes this point: "whatever else the Most Amazing Fact [i.e., the equivalence

instance, argued for the existence of his constants this way, appealing in part to diversity:

> The results quoted above [works by Einstein, Millikan, Warburg, etc], collected from the most varied branches of physics, present an overwhelming case for the existence of the quantum of action ... In view of all these results - a complete explanation would involve the inclusion of many more well-known names - an unbiased critic must recognize that the quantum of action is a universal physical constant, the value of which has been found from several very different phenomena to be [Planck's constant] (Planck, 1993, pp. 111–112).

For Planck, the confluence of the value of Planck's constant cannot be an accident (notice the undertone of *coding peculiarities* in the first alternative):

> Either the quantum of action was a fictional quantity, then the whole deduction of the radiation law was in the main illusory and represented nothing more than an empty non-significant play on formulae, or the derivation of the radiation law was based on a sound physical conception. In this case the quantum of action must play a fundamental role in physics.

For a very recent example, and rather amusingly, the position paper Huh et al. (2024) at the 41st International Conference on Machine Learning begins with the observation that "there is a growing similarity in how datapoints are represented in different neural network models ... which spans across different model architectures, training objectives, and even data modalities." In typical **Joint-carving** fashion, the authors claim that such representational convergence attests to an underlying joint to be carved: "our central hypothesis ... is that there is indeed an endpoint to this convergence and a principle that drives it: different models are all trying to arrive at a representation of reality." This statement is a clear echo of the

of definitions of computability] establishes, it shows that the notion which is stable under so many different formulations must be fundamental. "

sentiment expressed by Planck and others: the convergence of different models on the same representation is evidence that this representation is hitting upon something real.

The classic example from the history of science is the vindication of the existence of atoms and molecules (Perrin, 1916). According to Hudson (2020a), "what Perrin did was find the same value for Avogadro's number using a variety of experimental sources. These sources dealt with such diverse phenomena as the viscosity of gases, Brownian motion, critical opalescence, blackbody radiation, radioactivity, and other things." This is unmistakable confluence. For Perrin, the message was clear:

> Our wonder is aroused at the very remarkable agreement found between values derived from the consideration of such widely different phenomena. Seeing that not only is the same magnitude obtained by each method when the conditions under which it is applied are varied as much as possible, but that the numbers thus established also agree among themselves, without discrepancy, for all the methods employed, the real existence of the molecule is given a probability bordering on certainty. (Perrin, 1916, pp. 206–207; as cited in Hudson, 2020a)

Perrin's reasoning here has been the subject of much philosophical discussion. Proposed reconstructions[3] speak of inferring a common cause (Salmon, 2008), inference to the best cause (Cartwright, 1983), or evidence from robustness (Schupbach, 2018). The main theme is the same: although people disagree on what licensed Perrin to make that inference, or for that matter whether that inference was correct, there is no mistake that Perrin was making an argument by confluence, one that is strikingly similar to Kleene's argument by confluence.

Confluence, however, was not enough. In reflecting on the lessons drawn from this history, Maddy (2007) notes that already "a similar convergence of different methods for computing atomic weights emerged by 1860." (p. 404) And yet she suggests that "the evidence involved

[3]See Hudson (2020b) for a critical analysis.

in establishing the atomic hypothesis wasn't just more of the same, but a new type of evidence altogether." (p. 405) To wit, beyond just providing 13 different ways to achieve the same measurement of Avogadro's number, Perrin "traced 'the complicated nature of the trajectory of a granule', verified its 'complete irregularity', and confirmed Einstein's predictions." (ibid., p. 406) Hudson[4], who takes Perrin's vertical distribution experiment to have settled the case on behalf of atomic reality agrees:

> [Perrin's experimental approach] shows that he is doing much more than simply adding an extra method for deriving Avogadro's number, as though we can tip the scale on behalf of the correctness of this value by providing it with 13 as opposed to only 10 independent derivations. (Hudson, 2020b)

For Maddy's second-philosopher, what Perrin achieved was genuine "detection" of atoms. For Hudson:

> There is something uniquely special about Perrin's research into Brownian motion involving the vertical distribution experiment [beyond] ... merely affording the opportunity to interpret observed Brownian motion as if it were the product of molecular motion: in fact, this experiment involves the creation by Perrin of a model system in which we can observe in analogous fashion how the Brownian motion of particles ... produces the macroscopically observed relations between the height of an atmospheric gas and the density of this gas (i.e., the vertical distribution of a gas). This feature of the vertical distribution experiment distinguishes it as providing a unique form of evidence for the atomic hypothesis and explains the attitude of Perrin's contemporaries that Perrin's work on Brownian motion had in some sense resolved the debate about atomic reality. (ibid.)

[4]Hudson (2020b) goes as far as claiming that, contrary to popular belief, Perrin was not making an argument by confluence, but rather "reasoning analogically ... than robustly"

Notice how the timeline eerily resembles computability: confluence emerged already before 1936, but Gödel, for example, remained unconvinced until Turing's work appeared. This is the role of grounding in Turing's work: "The Turing Machine is not just another in the list of acceptable notions of computability - it is the grounding of all of them ... Once one has a grounding example in hand this changes - confluence now plays an epistemologically important evidentiary role." (Kennedy, 2020, p. 57).

At this point, may I suggest a leap of faith for the reader: *that we understand Perrin's detection of atoms and Turing's conceptual analysis of computability as on a par with each other*, that is, they play entirely analogous roles in **Joint-carving**, as special events that render evidence from confluence more compelling. Indeed, just like Perrin's experimental works constituted direct detection, Turing's own analysis of informal computability made "a direct appeal to intuition" (Copeland, 2024), using it so to speak as an (thought-)experimental tool to see what follows from first principles[5]. However, I should clarify that I am not claiming that their claims are subject to the same standards of acceptance and refutation or that their works reveal the reality of the world in similar ways or what have you. My claim is that, in their justificatory roles in conjunction with confluence, they serve surprisingly similar purposes. The similarity here is that both of these events directly exposed whatever previous results were converging upon: in the case of Perrin, all measurements returned Avogadro's number, because they were measuring the same things; and in the case of Turing, the definitions of computability all ended up equivalent, because they were touching upon the same idea.

Let us pause our exploration of the empirical landscape and take inventory of the souvenirs. We have seen that reading **Joint-carving**, i.e., that something has been done correctly, from confluence has been a salient practice of intellectual investigations. It has been a considerable driving force for the more empirical such. Of course, confluence is only a

[5]Curiously, Perrin saw himself as "show[ing] ... [the atomic hypothesis] is logically suggested by this phenomenon [of Brownian motion] alone" (Perrin, 1916, p. 7), an approach he considered "preferable".

part of the larger web of practice and beliefs in the empirical sciences, but hopefully the point is clear by now: similar kinds of justificatory forces of confluence are at play in the empirical sciences and in the analysis of computability (and other examples we have seen). With Perrin's example, I hope to have shown that (to play along with the present metaphor) that joint-finding, i.e., the kind of things that convince us that have hit the right idea, is an equally important aspect to **Joint-carving**. Bearing in mind that joint-finding in this context is a much more general phenomenon than Turing's grounding of computability, we now return to the safe havens of pure mathematics and observe justification in its natural habitat: axiomatic set theory.

A goal of modern (philosophy of) set theory is to understand the (meta-)mathematical facts surrounding axiom candidates: their consequences, consistency strengths, relations to better-known, canonical axioms, etc. In this regard piece of claimed evidence for the Continuum Hypothesis takes on the flavor of the kind of **Joint-carving** that we are considering here. Having defined a strong abstract logic based on considerations of generic invariance, Woodin puts his Ω-logic[6] to work. To wit, Koellner (2013) introduces the following theorem as a first step for making the case for $\neg\mathsf{CH}$:

Theorem 3.1 (Woodin)**.** *Assume that there is a proper class of Woodin cardinals and that the Strong Ω Conjecture*[7] *holds. Then*

1. *there is an axiom A such that*
   (a) *$\mathsf{ZFC} + A$ is Ω-satisfiable and*
   (b) *$\mathsf{ZFC} + A$ is Ω-complete for the structure $H(\omega_2)$*

[6]See Bagaria et al. (2006) for a technical introduction. For philosophical implications, see Koellner (2013) and Koellner and Woodin (2009)

[7]Definitions of the notions invovled here can be found in Bagaria et al. (2006) and Koellner and Woodin (2009). The latter also contains an instance of **Conjecture Heuristic** akin to sentiments about P vs. NP: "[the Ω-satisfiability of the Ω-conjecture] is a Σ_2-statement and there are no known examples of Σ_2-statements that are provably absolute and not settled by large cardinals. So it is reasonable to expect [it] to be settled by large cardinal axioms.

2. *Any such axiom A has the feature that $\mathsf{ZFC} + A \models$ "$H(\omega_2) \models \neg\mathsf{CH}$"*

For such an axiom A, write T_A for the theory of $H(\omega_2)$ as computed by $\mathsf{ZFC}+A$ in Ω-logic. Koellner summarizes the situation as "under these assumptions, there is a 'good' theory [of $H(\omega_2)$, in the form T_A] and all 'good' theories imply $\neg\mathsf{CH}$." Koellner points out a lingering worry: under the same assumptions, there are many such axioms and many such 'good' theories that are not identical (i.e., there are many axiom candidates A, B with $T_A \neq T_B$). For the case against CH, this "raises the issue as to how one is to select from among these theories." This worry is at least somewhat allayed by Woodin's identification of the axiom $(*)$ ("star").

Modulo large cardinals, the axiom $(*)$ itself exhibits an amazing amount of counfluence[8]. In Woodin (2010, Definition 5.1) it was originally formulated in modern set-theoretic language, as the conjunction of $\mathsf{AD}^{L(\mathbb{R})}$ and $L(\mathcal{P}(\omega_1))$ being a $\mathbb{P}_{\max}$-generic extension of $L(\mathbb{R})$. Viale, in light of the developments of Asperó and Schindler (2021) that showed it to be equivalent to a kind of forcing axiom, construed it as an axiom of "algebraic maximality," assertions that power sets are closed under various operations (2023, p. 11 and Theorem 9.10). Setting these equivalents aside, it is the following characterization that makes $(*)$ stand out in the present context:

Theorem 3.2 (Theorem 5.3 in Koellner (2013))**.** *Assume ZFC and that there is a proper class of Woodin cardinals. Then the following are equivalent:*

1. $(*)$

2. *For each Π_2-sentence φ in the language for the structure*

$$(H(\omega_2), \in, I_{\mathrm{NS}}, A \mid A \in \mathcal{P}(\mathbb{R}) \cap L(\mathbb{R}))$$

[8] This in itself subject the axiom to the rule of thumb of *diversity*, cf. Maddy (1988a), referenced in the beginning of Chapter 2

if

$$\mathsf{ZFC} + \text{“}(H(\omega_2), \in, I_{\text{NS}}, A \mid A \in \mathcal{P}(\mathbb{R}) \cap L(\mathbb{R})) \vDash \varphi\text{”}$$

is Ω*-consistent, then*

$$(H(\omega_2), \in, I_{\text{NS}}, A \mid A \in \mathcal{P}(\mathbb{R}) \cap L(\mathbb{R})) \vDash \varphi$$

With this property, the theory $T_{(*)}$ maximizes the Π_2 theory of the structure $(H(\omega_2), \in, I_{\text{NS}}, A \mid A \in \mathcal{P}(\mathbb{R}) \cap L(\mathbb{R}))$, something that is deemed desirable in set-theoretic (Maddy, 1998) and other mathematical/scientific (Baker, 2002) practices. The case of $\neg\mathsf{CH}$ is made, by first observing that $T_{(*)}$ is a good theory, and then via this maximum property (and appealing perhaps to diversity), showing that the axiom $(*)$ is desirable.

I take it that the rhetorical situation is quite analogous to the case of computability. On the one hand, we have a number of non-identical candidates: Turing machines and λ-calculus differ in their, say, conceptual naturalness or degree of dependence on the formalism; the T_A and T_B's in Koellner's examplme disagree on what is true in $H(\omega_2)$. On the other hand, all of the candidates are equivalent in the relevant aspects (capturing the same class of functions, providing a complete picture of $H(\omega_2)$ and settling CH). Joint-finding manifests differently in each case: we have seen that giving a formalism-free conception is one way to achieve this, as Turing did, but options vary, as we have seen in the case of Perrin and Koellner-Woodin. At any rate, what is common to each case is that some discovery takes place so that, to put Kennedy's characterization of grounding in more general terms, among the class of distinct candidates, one of them stands out as being indubitably adequate - as being right.

Of course, the question is still wide open whether these arguments surrounding the Continuum Hypothesis are successful[9]. I am not here to argue this matter one way or another,

[9]Especially in the case of CH, since there is an entirely analogous argument in its favor, appealing to the same considerations of Ω-completeness and maximality. See Koellner (2013) and Koellner and Woodin (2009).

but merely pointing out the similarity in the claims of justification and the intricate *pas de trois* between confluence, **Joint-carving**, and joint-finding. What this really shows, perhaps more convincingly, is that joint-finding is hard, and few and far between.

Admittedly, in the above I have allowed myself plenty of leeway in considering how these situations are alike. In my defense, again, I do not intend to claim that these arguments from confluence all carry the same weight or that they are subject to the same standards of confirmation and refutation. My goal has been to map out the various ways that one justifies something with confluence. In the face of overwhelmingly similar claims of **Joint-carving** in the presence confluence, from diverse fields of intellectual investigation, one cannot help but think that similar justificatory forces are at work across these examples.

For an explicit claim of a failed **Joint-carving** due to unsuccessful joint-finding, consider the case of algorithmic randomness. Celebrated results in this area speak of the equivalence between three independently motivated paradigms (via probability, betting strategies, and incompressibility) of characterizing what it means for an infinite binary sequence to be algorithmically random[10]. So celebrated, in fact, that Li and Vitányi utter another characteristic instance of (*Conf*):

> The fact that such different effective formalizations of infinite random sequences turn out to define the same mathematical object constitutes evidence that our intuitive notion of infinite sequences that are effectively random coincides with the precise notion of Martin-Löf random infinite sequences. (Li & Vitányi, 2008, pp. 221–222; as cited in Porter, 2021)

It cannot be an accident that independently motivated attempts at capturing randomness end up equivalent. Unsurprisingly, this invites speculation that there is something correct

[10]The reader is referred to Li and Vitányi (2008), Nies (2012), or Downey and Hirschfeldt (2010) for background details.

about these attempts. This facet of **Joint-carving** is particularly salient in the opening of Delahaye (1993), putting forward a Thesis:

> The definition of random sequences by Martin-Löf in 1965 and the other works on the so-called 'algorithmic theory of information' by Kolmogorof, Chaitin, Schnorr and Levin (among others) may be understood as the formulation of ... the Martin-Löf-Chaitin's Thesis: the intuitive informal concept of random sequences ... is satisfactorily defined by the notion of Martin-Löf-Chaitin random sequences.

In his treatment of the tenability of a Church-Turing-type thesis arising from the supposed confluence of numerous definitions of algorithmic randomness, Porter (2021)[11] called this phenomenon *Equivalence as Evidence Capturing* (the EEC claim), what we have been calling **Joint-carving**:

(the EEC claim) The equivalence of definitions of computable function provides evidence for the claim that these definitions capture the intuitive notion of an effectively calculable function.

Nevertheless, as Delahaye and Porter both agree, the situation with the proposed Martin-Löf-Chaitin Thesis suffers from additional complications that render it less convincing. A main difference in the case algorithmic randomness is whether the formal analysis by Martin-Löf and others really carves nature at its joint; or for that matter, whether there is a joint to be carved in the first place. For example, Porter maintains that

> There are no alternative definitions of computable function that are serious candidates for capturing the intuitive notion of computable function, a situation one

[11]Porter (2021) also documents a large number of occurrences of ((*Conf*)) in the algorithmic randomness literature. The reader is strongly encouraged to take a look and see how similar those are to the ones we have considered here.

> might describe by saying there is a unique locus of equivalent definitions of computable function (I shall sometimes also refer to this as a *locus of definitional equivalence*).

In our current metaphor, Porter is attributing the untenabiliity of the Martin-Löf-Chaitin Thesis to a lack of joint-finding in the case of algorithmic randomness:

> By contrast, this uniqueness datum is not present in the theory of algorithmic randomness: not every reasonable attempt to formalize the intuitive notion of random sequence has led to the same collection of sequences. [...] There are multiple definitions D of randomness such that (1) D-randomness is not equivalent to Martin-Löf randomness, and (2) D is also a locus of equivalent definitions of randomness. [...] Appealing to equivalence results as evidence for the claim that a formal definition captures some informal notion is convincing only in the presence of a unique locus of equivalent definitions.

Porter's final assessment of the supposed Martin-Löf-Chaitin thesis is that it tries to carve a joint where it ends up smashing a shin:

> Randomness-theoretic equivalence results suggest that the original target of definitions of randomness, the so-called intuitive conception of random sequence, is not a single conception that can be captured by any one definition of randomness, but rather, multiple definitions of randomness play a role in illuminating the concept of randomness.

So it would seem Porter's No-Thesis argument rules out the viability of the provisional Martin-Löf-Chaitin thesis, on grounds that the equivalence of ML-randomness, Kolmogorov-randomness, and martingale-randomness is just that: a mathematical fact that some things are co-extensional. In the presence of other plausible candidate explications of *the* intuitive concept of randomness, many of which are inequivalent to (say) ML-randomness, the significance of the foregoing equivalence is brought into question. In the current metaphor, Porter is charging the provisional Martin-Löf-Chaitin-Kolmogorov thesis with a lack of joint-finding.

If Porter's No-Thesis argument holds, then a thesis to the effect that ML-randomness is the correct formalization of randomness cannot obtain. Nevertheless, there is something to be said about the mere fact that independently motivated formalizations confluenced on something. Even though (granting Porter's argument) it is not *the* class of random sequences, it still enjoys a privileged status in practice[12]. Furthermore, the alternative candidates for formalizing algorithmic randomness all exhibit their own confluence phenomena. For example, with the necessary relativizations in place, 2-randomness can be equivalently characterized in terms of ML-tests, martingales, and Kolmogorov complexity.

Porter calls this schematic equivalence, meaning that these formalizations follow roughly the same schema. In each of the concrete instance of the schema, one finds the three usual definitions (modulo the relevant modifications) being equivalent again. For Porter, the significance of schematic equivalence is this:

> Although we do not arrive at a single extension of random sequences as capturing the intuitive conception of randomness, the schematic approach to the definitions of randomness provides the plurality of definitions of randomness with conceptual unity and gives a rationale for certain ongoing investigations being carried out by researchers in algorithmic randomness

In other words, although these clusters of little local confluences in the case of algorithmic randomness fail to convince us that (say) Martin-Löf-randomness managed to carve out a class of sequences that exactly captures our intuitive notion of an algorithmically random

[12] By this I mean it is covered in most texts, found in most research articles in this area, etc

sequence, the mere fact that equivalence abounds is already enough motivation (and has provided an array of tools) for researchers to launch a more serious study into this class and its relatives. The intellectual impulse induced by having observed these curious equivalences is the subject of the next section.

3.2 Remarkable Coincidence

The key component in Porter's No-Thesis argument is the lack of successful joint-finding. In his context, this refers to a lack of some motivating story that provides, in his terms, a locus of the equivalences. Competing candidates of randomness all receive the same kind of justificatory support from the respective groups of equivalence results, if any of them does. For this reason, **Joint-carving** fails to take place.

And yet we have the strong feeling that something is underlying these equivalences. This is again the feeling of *no accident*, that cries for an explanation. It will presently benefit us to think about what kind of explanation can or should be asked for.

In her discussion of mathematics in application, Maddy (2007, Chapter IV.2) challenges the common assumption that the so-called miracle of unreasonable effectiveness of mathematics in the sciences is a mystery that needs grounding. Confluence has certainly contributed to the impressiveness of the latter. Maddy cites Feynman: "There is a most remarkable coincidence: The equations for many different physical situations have exactly the same appearance. Of course, the symbols may be different - one letter is substituted for another - but the mathematical form of the equations is the same." (Feynman et al., 1964; as cited in Maddy, 2011, p. 333)

This type of **Remarkable Coincidence** will be the subject of this section, but let us first stick with Maddy and Feynman. Faced with overwhelming confluence of the equations, one is naturally tempted to think that there is something behind it. Maddy goes on to cite Feynman suggesting that, despite that urge to find an explanation, there is no principled reason why all these different physical situations should be governed by the same equations, least of all some sort of "underlying unity of nature." She presents an abundance of evidence, ranging from history of mathematics to the statistics of coincidences to our cognitive whims and fancies, eventually arriving at the conclusion that the miracle of applied mathematics is not so miraculous after all: much of the feeling of remarkability can be explained away by looking at careful examination of the history of the mathematics involved and the nebulous criteria of what counts as successful application. Ultimately, she attributes the sense of miracle to our susceptibility to apparent coincidences as pattern-seeking creatures.

I fully agree with Maddy here, but this is not the place to do it justice. The applicability of mathematics is a fascinating question worthy of a more focused treatment. Given that it is somewhat distant from our current concerns, I am only referencing it to sketch a useful parallel in the argumentative strategy that I am going to adopt. Although her concerns above are somewhat distant from ours, in the treatment of **Remarkable Coincidence**, I shall adopt a spirit similar to Maddy, namely that there is often no need to read too much into a coincidence, no need to seek some sort of principled grounding for why confluence occurs. (In cheekier language: joint-finding is so overrated.) I will differ with Maddy in one key respect, however. Whereas Maddy's concerns about applied mathematics lead her to explain away the appearance of miracles, I shall fully embrace these miracles and maintain that our susceptibility to apparent coincidences as pattern-seeking creatures is a good thing. Or at least it is useful in doing actual mathematics.

Whether there is an underlying unity of nature that subjects it to mathematical treatment or the miracle of mathematics ultimately turns on an illusion whipped up by our cognitive predilections, I take it that the kind of remarkable coincidences that Feynman alluded to has also been remarkably fruitful, both in presenting new ways to apply mathematics in the sciences and in philosophical reflections about the nature of such coincidences. My point is that, suspending judgment on the idea of successful capture or accurate description or what have you, **Remarkable Coincidence** gestures not so much towards something correct, but rather something interesting and fruitful in the first place.

My reason for thinking so comes again from observations of mathematical practice. Recall a sentiment expressed in Epstein and Carnielli (2008) in the previous section, this time quoted fully to illustrate this kind of low-commitment attitude towards confluence:

> All the authors we have quoted above are in agreement on one thing: [Church's Thesis] is not part of mathematics, it is not part of the theory of recursive functions or Turing machines. Those theories are interesting in their own right even if Church's thesis/definition were to be abandoned. Whatever else the Most Amazing Fact [the equivalence of definitions] establishes, it shows that the notion which is stable under so many different formulations must be fundamental.

Here, Church's Thesis is being understood as a principle relating the formal and the informal. Note that this does not mean just **Rigor Assurance** (the paragraph after the quoted passage proceeds to argue against using Church's Thesis this way, an argument that we touched upon in Sections 2.3 and 2.3). In the present framework, Epstein and Carnielli are casting out from mathematics this supposed philosophical commitment in **Joint-carving** entailed by accepting Church's Thesis. Regardless of whether one thinks the definitions of computability are *really* about the One True Notion, one still has a large body of autonomous mathematical facts in favor of studying the theory of recursive functions or Turing machines and whatnot. This is a first evidence that **Joint-carving** and **Remarkable Coincidence** can be teased apart. We shall see that many instances of confluence can be pursued fruitfully without presupposing that they carve joints.

One such instance of **Remarkable Coincidence**, one reminiscent of the facts that motivated Martin's Conjecture, is found in set theory[13]:

> What's emerged over the years is that many theories set theorists consider turn out to be equiconsistent with ZFC extended by one large cardinal axiom or another. Moreover, these large cardinal axioms are linearly ordered by their consistency strength. Of course it's possible to concoct a theory for which this fails, but as a straightforward matter of empirical fact, it has been true for 'natural' theories entertained to date. (Maddy and Meadows, 2020)

The justification of the determinacy hypotheses (Maddy, 1988b) marks another celebrated chapter in this research. A pleasant outcome is that determinacy hypotheses not only provide solutions to insurmountable problems left open by the Polish and Moschow schools of mathematicians, they are also deeply connected with the large cardinals. Remarking on the equiconsistency between the hypothesis $\mathsf{AD}^{L(\mathbb{R})}$ and the existence of infinitely many Woodin cardinals, Maddy writes:

> Considering that determinacy and large cardinals arose in the course of such disparate, apparently unrelated contexts of mathematical inquiry, this ultimate equivalence is quite surprising and impressive: 'This sort of convergence of conceptually distinct domains is striking and unlikely to be an accident' (Koellner, 2006, p. 174). Our second-philosophical Objectivist understands the situation this way: the fact that two apparently fruitful mathematical themes turn out to coincide makes it all the more likely that they're tracking a genuine strain of mathematical depth. (Maddy, 2011, p. 129)

[13] I should note that bits of **Coding Invariance** are at play here too. Since these theories are c.e. extensions of ZFC, they are subject to the manner in which they are enumerated. Hamkins (2022) takes advantages of this and devises methods to enumerate theories having the same sentences but different consistency strengths (this is a lot like the coding peculiarities we saw in Turing's completeness theorem in that section). Hamkins takes this as evidence that there are intensional aspects to consistency strength comparisons. I leave this issue open here.

In fact, Koellner (2006) sees much value in this kind of confluence. Koellner describes the convergence between classical regularity properties, determinacy hypotheses, and inner models for large cardinals as "striking" and that the equivalence between the latter two "evidence that both are on the right track."

The common interpretation of this "right track" is that these are likely the right axiom candidates. Such is **Joint-carving**, and it makes one wonder if there is any motivating story that can be told. But, following the spirit I have sketched above, I would like to suggest that, in this case, a story is not needed. Their remarkability remains the same whether or not we understand them in the context of axiom justification. In other words, the confluence we have seen is definitely gesturing towards something, and one can follow this clue fruitfully without any joint-finding: disregarding the issues of whether an axiom is right or wrong, there is still the undeniable fact that this theme has been pursued fruitfully and remains a stable motivator for new set-theoretic research. In sum, my suggestion is to take Maddy's second-philosophical Objectivist literally: "the fact that two apparently fruitful mathematical themes turn out to coincide makes it all the more likely that they're tracking a genuine strain of mathematical depth."

The very next sentence of Feynman et al. (1964), right after where Maddy cuts off Feynman's quotation, clarifies the view further: "This [having seen the remarkable coincidences] means that having studied one subject, we immediately have a great deal of direct and precise knowledge about the solutions of the equations of another." The observation is straightforward: in many cases we are not in the business of carving joints, but rather in the business of producing mathematical results. The confluence of definitions is a strong motivator for further research, regardless of whether we have a target or explanation in mind. This is the spirit of **Remarkable Coincidence**. It concerns the justification of a particular pursuit or a way of doing things, without much thought about whether there is a privileged or correct way of doing things. By way of comparison, and to once again put things in slogans, **Remarkable Coincidence** may be best thought of as **Joint-carving** suppressing any background notion of being correct or privileged. For this reason I consider it to be low in philosophical commitment: one need not be bothered with a pre-theoretic notion, or any criteria of naturalness or correctness. To give it a more informal spin: **Joint-carving** tells us something has been done correctly, **Remarkable Coincidence** tells us something can be done (or has been done) fruitfully.

In a somewhat distant field, we also see intricate dynamics between **Joint-carving** and **Remarkable Coincidence** in the study of dynamical systems. I consider this next example to be particularly strong evidence that these two facets of justification by confluence can be teased apart. For this purpose, the reader is encouraged to keep in mind the development of the Church-Turing Thesis for parallel comparison throughout this next example. First, some definitions.

Let **MPT** be the collection of invertible measure-preserving transformations defined on $[0, 1]$. A Polish topology can be induced on the space, using a relatively natural metric. Let $SIM([0, 1)^{\mathbb{Z}})$ denote the space consisting of the collection of standard Borel probability measures on $[0, 1)^{\mathbb{Z}}$ that are invariant under the shift map $sh(f)(n) = f(n+1)$; endowed with the weak* topology, it too forms a Polish space. Under a suitable map $\pi : SIM([0, 1)^{\mathbb{Z}}) \rightarrow \mathbf{MPT}$, elements of $SIM([0, 1)^{\mathbb{Z}})$ can be viewed as codes for measure-preserving transformations. Finally, a class P of measure preserving transformations is called a dynamical property iff for all pairs (S, T) of isomorphic transformations, $S \in P$ iff $T \in P$[14].

The central motivation of Foreman (2010) derives from the following seminal theorem proved in the 90s (Glasner & King, 1998; Rudolph, 1998).

Theorem 3.3 (Glasner and King, Rudolph)**.** *Let P be a dynamical property that has the property of Baire. Then P is generic in* ***MPT*** *iff P is generic in $SIM([0, 1)^{\mathbb{Z}})$.*

[14]The reader is referred to Foreman (2010) and references therein for further technical background.

This is indeed a remarkable coincidence. For context, several lines of research in dynamical systems concern the classifications and/or complexity of certain classes of measure-preserving transformations. To do this formally, one needs to consider collections of such transformations as topological spaces, thus endowing them with, e.g., a Borel structure[15]. The preceding theorem shows that two natural candidates for doing so end up being equivalent in a special sense. This and related results motivated Foreman to formulate the following thesis.

(Rudolph's Thesis). Any two models for the measure preserving transformations are equivalent and have the same generic dynamical properties.

After making the usual disclaimers that the thesis is an informal statement that is subject to proof only after the terms are made precise (compare again with the Church-Turing Thesis), Foreman proceeds to lay out his goals: "to make this thesis precise in a plausible way and verify it to the extent possible". He does this by first carefully defining what "equivalent" means and what "model" means. For instance:

Definition 3.4 (Definition 10 in Foreman (2010))**.** *A model is a pair* (X, π) *where* X *is a Polish space and* $\pi : X \to \mathbf{MPT}$ *is a continuous function for which there is a comeager set* $A \subseteq \mathbf{MPT}$ *such that for all* $T \in A$*,* $\{x \in X \mid \pi(x) \cong T\}$ *is dense in* X*.*

Obviously this is not the place to go into technical details. The above example is meant to highlight, and this is the key observation for our purpose here, that this definition is almost entirely motivated by the mathematics, rather than intuitive reflections on what pre-theoretic notions they are supposed to mean or capture: the map π (in the case of $\pi : SIM([0,1)^{\mathbb{Z}}) \to \mathbf{MPT}$ as well as the general case) ought to be continuous as opposed to the more natural Borel measurable, because would render the definition too weak, allowing simple counterexamples; the added clause of the existence of a comeager set is due to the Rokhlin's lemma, which places more stringent control on the topology in order to "prevents

[15] This is the sort of settings contemplated by Gao's Thesis.

the existence of exotic open sets" (or an instance of what I have called *coding peculiarities* previously); and elsewhere, it is remarked that taking "equivalent" to mean "homeomorphic" would be too strong, again giving rise to easy counterexamples, and so genericity is brought into the fold, because describing which properties are generic has been a "fruitful area of study in ergodic theory."

The example of Foreman (2010) serves to illustrate, in my view, a paramount case of **Remarkable Coincidence**, that is, **Joint-carving** suppressing any background notion of being correct or privileged, guided only by what the mathematics tells us. Quite markedly, it is a distinct kind of exercise, in contrast to Turing's thorough reflection on the ability of the idealized human computor, thus motivating the definition of Turing machines with his musings. Foreman was obviously not concerned with motivating a notion of "model" with some careful philosophical reflection, the way Turing did with "computable". That was never the goal. After all, why should a remarkable mathematical coincidence need to answer to extra-mathematical, pre-theoretic tribunals, to borrow a phrase from Maddy (1997)? The confluence is a strong indication for fruitful research, regardless of whether we have a target or explanation in mind.

In the absence of a clear pre-theoretic notion, Foreman can only be understood as following the mathematical clues left by confluence and trying to pursue it as far as possible. This attitude is perfectly summarized by Maddy's second-philosophical Objectivist:

> Once we have a concept that's mathematically fruitful, it's a rational policy to exploit it further, to try to extend it in ways that seem "natural" or harmonious with its leading intuitions ... What's striking is that all these perfectly reasonable ways of proceeding are in fact grounded in their promise of leading to the realization of more of our mathematical goals, to the discovery of more fruitful concepts and theories, to the production of more deep mathematics ... What does matter,

all that really matters, is the fruitfulness and promise of the mathematics itself. (Maddy, 2011, pp. 136–137)

With this in mind, consider Porter's takeaway from the confluence phenomena in randomness. The main obstacle of an analogous Church-Turing thesis was that the confluence of equivalence definitions fails to carve any sensible joint. In the previous section, we saw Porter rule out the viability of a Church-Turing Thesis for algorithmic randomness, on grounds that "the so-called intuitive conception of random sequence is not a single conception." Nevertheless, the sheer equivalence of natural, independently motivated definitions itself provided sufficient motivation, and technical tools, to further study these definitions (and their equivalence).

Porter (2021) spends considerable amount of space analyzing the significance of the equivalences, even after establishing that they cannot possibly be of the **Joint-carving** kind. Our present framework shows that whatever else they suggest, the equivalences show that pursuing schematic equivalences for variant notions of algorithmic randomness is likely fruitful.

3.3 Counterexample Resistance

The final facet of justification by confluence, one embodied in Kleene's argument by confluence, begins with an anecdote. According to his personal recollection of the development, when Kleene first heard Church's Thesis, his initial reaction was definitely skeptical: "he can't be right." But he became an instant convert when he realized he could not diagonalize out of the class of λ-definable functions (Crossley, 2006, p. 7). On top of Turing's conceptual analysis and the confluence of definitions, Kleene was further convinced by how the concept seemed resistant to the kinds of cheeky balloon-popping trickeries that logicians have prided themselves on in other cases. This **Counterexample Resistance** led him to formulate the following as its own freestanding evidence for the Church-Turing Thesis:

> The exploration of various methods which might be expected to lead to a function outside the class of the general recursive functions has in every case shown either that the method does not actually lead outside or that the new function obtained cannot be considered as effectively defined, i.e. its definition provides no effective process of calculation. (Kleene, 1952, pp. 319–320)

For what it's worth, Kleene had every right to doubt Church's Thesis. After all, diagonal arguments have been such a reliable go-to for the logicians, successfully applied time and time again whenever a result seems too good to be true. Gödel, for example, called the resistance to diagonalization a miracle: "For the concept of computability ... [b]y a kind of miracle it is not necessary to distinguish orders, and the diagonal procedure does not lead outside the defined notion." (Gödel, 2001; as cited in Kennedy, 2020, p. 76)

In the above, Gödel was drawing a contrast with definability and provability, both susceptible to diagonal arguments. It turns out that the **Counterexample Resistance** of computability manifests itself in a much broader fashion than just being undiagonalizable. As a first hint, we have Rogers noticing that adding additional computational resources to Turing machines does not really change the computational power:

> These equivalence demonstrations can be generalized to show that over certain very broad families of enlargements of these formal characterizations the class of partial functions obtained remains unchanged ... For example, if we allow more than one tape, or other symbols than 1 and B, in the definition of Turing machine, the partial functions obtainable are still partial recursive functions. (Rogers, 1987, p. 18)

To be fair, Rogers is touching upon **Coding Invariance** here, but implicit in this example is the "enlargement" aspect of it, which this section intends to draw out. We have seen that the "dishonest" codings in the sense of Boker and Dershowitz might accidentally render the uncomputable computable. But there have been historically actual instances where this was the express purpose, i.e., to enhance the power of our definition and see what happens. Notwithstanding, the class of computable functions has remained surprisingly robust. We will see a number of such instances below. The reader is especially encouraged to keep Rósza Péter's example close in mind, as it will return in the discussion of a surprisingly analogous situation in constructible universes from extended logics.

In the early days of computablity, Rósza Péter[16] set out to determine the power of certain apparently stronger formulations of computability, by allowing more computational resources. These include what we call today course-of-value recursion ("*die Wertverlaufs-rekursion*") and nested recursion ("*die eingeschachtelte Rekursion*"). Péter eventually met the same plot twist as Kleene[17]:

> One might think that the recursive functions using different choices of recursion concepts generate different classes of number-theoretic functions, the narrowest of which is when only primitive recursion is allowed (trivially, the other recursions contain this case as a special case). I will show however, that adding [course-of-value recursion] and [nested recursion] to the class of primitive recursive functions does not extend that class. (Péter, 1935, p. 615)

The theme addressed by Rogers and Péter, i.e. equipping the concept at hand with more and see what happens, has been a constant thread for the better part of the development of

[16]To whom the term *primitive recursive* (*primitive Rekursion*) is attributed (Gandy, 1988, p. 68).

[17]Péter (1935, p. 615): "Man könnte glauben, dass die rekursiven Funktionen bei verschiedener Wahl des Rekursionsbegriffs verschiedene Klassen von zahlentheoretischen Funktionen bilden; die engste, wenn nur die primitive Rekursion zugelassen wird (es ist trivial, dass die übrigen Rekursionen diese als Spezialfall enthalten). Ich werde aber beweisen, dass die Hinzunahme der Rekursionen 1. [die Wertverlaufsrekursion] und 2. [die eingeschachtelte Rekursion] die Klasse der durch primitiven Rekursionen definierten rekursiven Funktionen nicht erweitert." Translation my own, with helpful suggestions from Chris Mitsch.

computability theory. For a more recent example, recall the paper by Leeuw et al., whose declaration of informality was cited as an example of **Rigor Assurance** earlier. Its opening line: "the following question will be considered, in this paper: Is there anything that can be done by a machine with a random element but not by a deterministic machine?"

In more tangible terms, they considered the following enhancement of the classical Turing machines: the machine is equipped with a coin-flipping state, such that, in this state, the machine flips a coin and writes the result on the current cell. The main idea is to introduce a source of randomness to the process of computation. Suppose an uncomputable real $x \subseteq \omega$ is computable in this sense with probability 1, then x has a good chance of serving as a counterexample to the classical Church-Turing thesis, in that coin-flipping (or however one wants to precisify an action with random outcome) can be considered an effective action that can carried out mechanically.

Given the section title, the reader can rest assured that such a real x does not exist. This is independently proven by Sacks (1966) for hyperarithmetic reduction and Leeuw et al. (1956) for Turing reduction:

Theorem 3.5. *For any real a, if $\{x \mid a \leq_T x\}$ has positive measure, then a is computable.*

The philosophical import of the foregoing theorem, in service of preemting a counterexample to the Church-Turing Thesis, is clearly articulated by Downey and Hirschfeldt in their popular textbook on algorithmic randomness:

> ...[Theorem 3.5] has a very interesting corollary ... For a set A, let $A^{\leq_T} = \{B : A \leq_T B\}$. It is natural to ask whether there is a noncomputable set A such that $A^{\leq_T}$ has positive measure. Such a set would have a good claim to being "almost computable". Furthermore, the existence of a noncomputable set A and an i such that $\mu(\{B : \Phi_i^B = A\}) = 1$ could be considered a counterexample to the Church-Turing Thesis, since a machine with access to a random source would be able to

compute A. [The Sacks-LMSS theorem] lays such worries to rest. (Downey & Hirschfeldt, 2010, p. 358)

Theorem 3.5 is only a member in a family of deep theorems. The phenomenon generalizes to stronger notions of computability and also with genericity replacing randomness, and at the level of projective reducibility it is another reminder of the intricate interplay between determinacy hypotheses and the structural theory of the projective classes. This **Remarkable Coincidence** has been the motivation for a number of research papers, but I digress[18].

The preceding examples are meant to demonstrate a common pattern: putative enhancements that might lead outside of the class of computable functions do not in fact do so. They end up equivalent to the original definitions, despite appearing to be strengthenings. It is as if tipping the scales in favor of the uncomputable does not actually tip the scales at all. This robustness of computability under apparent strengthening illustrates the final facet of confluence that we will be considering.

Counterexample Resistance, in the sense considered above, appears in the curious case of generalized constructibility. For an abstract logic $\mathcal{L}$ let $C(\mathcal{L}) = \bigcup_{\alpha\in\mathrm{Ord}} L'_\alpha$, where

$$L'_0 = \emptyset$$

$$L'_{\alpha+1} = \mathcal{D}_{\mathcal{L}}(L'_\alpha)$$

$$L'_\lambda = \bigcup_{\alpha<\lambda} L'_\alpha, \text{ for limit } \lambda$$

where for any set M, $\mathcal{D}_{\mathcal{L}}(M) = \{X \subseteq M \mid X \text{ is a } \mathcal{L}\text{-definable subset of } M\}$. That is, $X \in \mathcal{D}_{\mathcal{L}}(M)$ iff X is of the form $\{x \in M \mid (M, \in) \vDash_{\mathcal{L}} \varphi(x, \vec{a})\}$ where φ is a formula in the logic $\mathcal{L}$ and $\vec{a}$ are parameters from M.

[18]The interested reader can find a treatment of this family of theorems, in relation to their proofs using metamathematical techniques, in Chapter 4.

Kennedy et al. (2021) proved the following remarkable theorem:

Theorem 3.6 (Kennedy, Magidor, Väänänen)**.** *Suppose $\mathcal{L}$ is $\mathsf{ZFC} + V = L$-absolute*[19] *with parameter from L and has $\mathsf{ZFC}+V = L$-absolute syntax with parameter from L, then $C(\mathcal{L}) = L$.*

Corollary 3.7. *$C(\mathcal{L}) = L$ if $\mathcal{L}$ is first-order logic equipped with any of the following.*

1. *cardinality quantifiers Q_α, $\alpha \in \mathrm{Ord}$*
2. *equivalence quantifiers Q^E_α (this quantifier has the meaning "φ defines an equivalence relation with $\geq \aleph_\alpha$-many equivalence classes")*
3. *recursive countable conjunctions and disjunctions*
4. *recursive game quantifiers: $\forall x_0 \exists x_1 \forall x_2 \exists x_3 ... \bigwedge_{n\in\omega} \varphi_n(x_0, y_0..., x_n, y_n)$*
5. *well-ordering quantifier*

This is characteristic confluence, and Kennedy takes this to be strong evidence that we have a Church-Turing-type thesis:

> We suggest that this manifests a remarkable independence of L from the formalism used, and in that sense provides evidence for Gödel's suggestion that constructibility might be a good candidate for a formalism independent notion of definability ... Constructibility being not particularly sensitive to the underlying logic in that sense gives evidence that a type of Church-Turing thesis holds for L, namely invariance with respect to a certain large class of logics. (Kennedy, 2013; reprinted in Kennedy, 2020)

Kennedy situated Theorem 3.6 and Corollary 3.7 in the context of answering Gödel's search for a candidate notion of definability, and also in her broader project exploring the

[19] This means that the syntax and semantics of $\mathcal{L}$ is $\mathsf{ZFC} + V = L$-provably Δ_1

nature of formalism-freeness and formalism-entanglement, using the range of logics as a test measure for such. Our concerns here are mostly orthogonal to that admirable project, so I shall remain silent about issues of formalism-freeness and to what extent this serves as an answer to Gödel's call for such a candidate. Nevertheless, seeing as this is a characteristic appeal to confluence, and having developed the framework in this dissertation, I believe we are at a position to contribute to the discussion.

In her review of Kennedy (2020), Maddy asks:

> "Constructability" in the work of Kennedy and her co-authors is the most straightforward example: the notion picked out by the usual first-order definition of L is shown to be more formalism-free than one might expect, though still to some extent logically entangled. But what exactly are we to conclude from these observations? (Maddy, 2021)

By now we have seen a number of cases where confluence is called upon to provide justification for something or other. If the taxonomy I have provided is worth our while, then one productive approach to answering Maddy's question is to consider the ways in which the confluence in the case of constructibility is similar to the examples we have considered throughout the example. The goal is that, by comparing the role of confluence played here to those played in the other boldfaced items, we will obtain a deeper understanding of which facets of justification by confluence are at play here. I shall proceed to do this now.

As it turns out, the situation with generalized constructibility is a surprising parallel to Péter's investigation of the primitive recursive functions, our primary example of **Counterexample Resistance**. In Péter's case, the additional computational resources of course-of-value recursion and nested recursion did not actually extend the class of primitive recursive functions. In Kennedy's case, the additional expressive resources of the various logics considered do not actually extend the class of constructible sets. Both enhance the underlying formalism in the same way, by permitting expressive apparatus seemingly beyond the first order. Both cases are resolved in the same way, too: Kennedy et al. showed that the satisfaction predicates of $\mathsf{ZFC} + V = L$-absolute logics for any particular level in their respective constructible hierarchy are always definable in some sufficiently high initial segment of L; thus letting L run long enough, it will eventually be able to encapsulate the additional, higher-order expressive resources. Péter, on the other hand, makes use of similar coding machinery to recover the additional expressive power of course-of-value recursion and nested recursion in the primitive recursive functions. All that is needed to be done in both cases is to let the "construction/computation" in question run long enough, and the additional expressive resources will be absorbed by the original formalism.

It is of interest, and perhaps even more surprising, to note that the same can be said about apparent weakenings of both cases too. Kunen (1980, p. 183) assigns the following exercise[20]: modify L's hierarchy such that at successor levels, only first-order definability *without* parameters is permitted; the resulting hierarchy is still L. In her paper above, Péter also remarked the same class of functions can be obtained by allowing the simplest form of primitive recursion, i.e., having no parameter. So in both cases, the underlying formalism exhibits strong resistance to both apparent strengthenings and weakenings.

This close parallel shows that the results of Kennedy et al. certainly gesture towards **Counterexample Resistance**. One may additionally point to the proliferation of recent inner model theory publications on these generalized constructible hierarchies as **Remarkable Coincidence**. But is there more? Maddy, for example, seems to doubt that **Joint-carving** is unlikely:

> If the case [of the equivalences of different generalized L's] is also to serve as an instance of variety-two formalism freeness, each of the elements of that calculus - the carefully calibrated list of stronger and stronger extensions of first-order

[20]The reader can find a solution sketch in Sorouri (2022)

logic - would have to be as conceptually distinct, as independently motivated, as the various definitions of "computability". This seems to me a stretch. (Maddy, 2021)

Here, by "variety-two formalism freeness", Maddy is describing the paradigmatic case of **Joint-carving** in computability: a number of independently motivated formalizations all converged, and in particular they all converged on a right idea grounded by Turing's analysis. In fact, the charge of "conceptually distinct" and "independently motivated" can be made precise via an elementary coding argument, from which we obtain[21]:

Proposition 3.8. *Suppose there is a definable (definable in the generalized logic) well-ordering of the universe for $C(\mathcal{L})$, then $C(\mathcal{L}) = L[A]$ for some class of ordinals A.*

So it would seem like Corollary 3.7 mainly showed that, for a number of classes A, the classes $L[A]$ are still L. This seems like a disanalogy from the case of **Joint-carving** in computability. To see why, it is illuminating to refer to the common pedagogical parlance, which talks of relativized constructibility using the analogy of oracles. We may think of structures like $L[A]$ as the constructible universes generated in the language $\{\in, A\}$, so that at each stage we are free to inquire about membership in A. This is often compared to equipping Turing machines with oracles (construed as sets of natural numbers), such that during the course of computation we may freely inquire about membership in the oracle. This analogy is attested, for example, in the Hamkins and Lewis (2000), in which it was used as a guiding principle to formulate what it means to have a set of reals as an oracle.

Taking this analogy seriously, we may look at what happens to Turing-computability if we were to add sets of natural numbers as oracles to Turing machines. For quite a few oracles (e.g, those that are computable to begin with), the collection of extended Turing-computable

[21] According to Kennedy et al. (2021), "most logics that one encounters in textbooks and research articles of logic" will result in a constructible universe with a definable well-ordering.

sets is coextensive with the collection of Turing-computable sets. But from this fact alone we are not compelled to accept the Church-Turing Thesis. So the framework developed in this dissertation provides further insight into Maddy's challenge: the confluence of various constructible hierarchies shows the **Counterexample Resistance** of iterated first-order definability, and it misses the mark of being the kind of **Joint-carving** that would be a strong argument for a Church-Turing Thesis, because the situation is more analogous to showing that various sets of natural numbers are computable and hence, as oracles, will expand the class of computable sets.

This does not mean, however, that **Joint-carving** is impossible for constructibility. I am merely making the claim that, under the general taxonomical framework of justification by confluence I have sketched, the kind of evidence characteristic of **Joint-carving** is not found in the results of Kennedy et al. (2021). For possible evidence that renders the situation more closely aligned with that of computability, I direct the reader's attention to the growing field of transfinite computation. A good point of reference can be found in the recent text Carl (2019). Two models of transfinite computation, ordinal register machines and ordinal Turing machines, generalize a normal Turing machine by allowing Ord-length computation, among other things. There, we have a familiar confluence result[22]:

Theorem 3.9 (Koepke)**.** *A set of ordinals is ordinal-Turing-computable iff it is ordinal register-computable iff it is in* L*.*

It is possible that L's confluence might be tracking some notion of transfinite computation. Of course, transfinite computation as presented above is a sufficiently formal notion. We may ask if this formal notion captures any informal notion. The obvious candidate is some kind of generalization of the pen-and-paper computability that Turingcomputability intends to capture. An alternative candidate is feasibility by an idealized agent in the foundation of

[22]Benjamin Seyfferth's doctoral dissertation has devised a generalization of λ-calculus on the ordinals, which is also shown to be equivalent to ordinal register machines and ordinal Turing machines.

mathematics. Reference to such an idealized agent can be found, for instance, in Takeuti (1985):

> Foundational problems begin when we realize that we cannot examine infinitely many objects one by one. However, it is very easy for us to imagine an infinite mind which can do so. Actually by working in mathematics we have been building up our intuition on what an infinite mind can do. An infinite mind must be able to operate on infinitely many objects as freely as we operate on finitely many objects.

I should refrain from saying more about this, lest we be side-tracked. The preliminary sketch of such a possibility is only meant to gesture towards a possible venue for **Joint-carving**. The main point is, just like Turing's motivating example of an idealized human computor, we might consider Takeuti's remark as motivating a notion of feasibility that manifests in such an imaginary infinite mind. Whether that notion is well-grounded, however, and whether that notion is something that's captured by transfinite computation, is an entirely different project. Philosophically, this motivates investigations towards the plausibility of an infinitary Church-Turing thesis, such as the one that Carl considered in the end of Carl (2019). We might also ask whether L is a good candidate for Takeuti's feasibility of an imagined infinite mind, and what philosophical import that thesis might have on the broader subject of philosophy of mathematics.

3.4 Conclusion

Jointly, Chapters 2 and 3 have appeared to be a sort of discursive survey of various instances when confluence results have been used to justify or support certain beliefs or practices. I have attempted to provide a taxonomy of these justificational purposes that confluence serves. I remain confident that the list is far from exhaustive. After all, almost all taxonomical attempts have been similarly incomplete at a first pass.

But I hope to have made plausible the central thesis I want to put forth: that appeals to confluence actually serve a wide variety of justificational purposes that are largely orthogonal to each other. We have seen all such purposes arise, in some fuzzy and conflated fashion, in the case of the Church-Turing Thesis. And the crucial finding here is just as I have forecasted in the beginning of the paper: that in many actual instances, appeals to the Church-Turing Thesis are really just appeals to one of these many justificational purposes served by confluence. We have also seen that the classification framework here finds broad applications to existing philosophical questions, and hopefully in some cases has led to a rather satisfactory resolution of them.

There remains lots to be done in this line of work. One may expand, refine, or otherwise challenge the taxonomy that I have sketched in this paper. Or one can apply this to other instances where confluence/equivalence/convergence results have been used to justify one thing or another. There are two areas ripe with these kinds of examples. One is the field of theory equivalence in the philosophy of science literature. The other is (philosophy of) category theory, where one has worked out a plethora of notions of equivalences, relating diverse mathematical structures in truly remarkable ways. Lamentably I have not touched upon here because my knowledge thereof is embarrassingly limited. I nevertheless hope that the kind of analyses provided in this paper will seem inviting and encouraging to those more knowledgeable in these areas.

Chapter 4

Metamathematical Methods in Descriptive Set Theory: a Case Study in the Taxonomy of Proofs

4.1 Overview

The focus of this chapter is the use of metamathematical methods in descriptive set theory. My usage of the term “metamathematical methods” is meant to elicit analogous connotations of terms like “combinatorial methods”, or “topological methods”, namely proofs or techniques that make use of theorems or tools from a particular branch of mathematics. Admittedly, every proof in metamathematics, for instance proofs of Gödel’s incompleteness theorem or the independence of the continuum hypothesis, uses metamathematical methods (because we are dealing in metamathematics after all!), so I must first clarify what I mean.

By metamathematical methods[1], I mean the collection of concepts, tools, and techniques

[1]For some of the history of these methods, see Kanamori’s account of the works of Levy (Kanamori, 2006),

originating in the study of metamathematics: these include forcing, inner models, (countable transitive) models of set theory, absoluteness, as well as considerations of consistency, satisfaction, provability, and logical complexity. Tentatively, I would like to consider proofs that use these methods, but are themselves proofs of statements in fields relatively distant from metamathematics[2]. Rather simply put, I would like to consider proofs of mathematical theorems using metamathematical methods. In what follows, terms like "metamathematical proof" or "metamathematical methods" will be construed as proofs and methods in this context. Descriptive set theory, as we shall see, provides a fertile ground for analyzing such examples and their interesting philosophical offshoots.

This chapter has a two-fold purpose: on the practical side, to provide a rough collection and classification of the various types of proofs in descriptive set theory that can be characterized by their use of metamathematical methods. In doing so, I also aim to track the degree of involvement of metamathematical tools in these proofs, and hope to demonstrate that there is a sui generis methodology, in full analogy with, say, "algebraic methods". And on the philosophical side, the goal is to probe perhaps more general questions about the nature proofs and methods, particularly regarding what it means to characterize a proof by the method used, and what a method imports in a proof.

Since I am working towards a peculiar goal, I will have to proceed somewhat peculiarly. This chapter will be structured as follows: to begin, I will first motivate this project a little by relating it to a broader philosophical project born out of Dawson's monograph (Dawson Jr., 2015) about the values of having multiple proofs. Next, I will consider an obvious non-example of metamathematical proof (that makes use of metamathematical methods), which

who " played an important role in raising the level of set-theoretic investigation through metamathematical means to a new height of sophistication." See also Avigad (2010) for more on the metamathematical tradition.

[2]This is not to imply that I am committed to a serious distinction between what is mathematics and what is metamathmatics. I am merely reflecting usage in the literature, for instance see Raisonnier's conspicuously titled *A mathematical proof of S. Shelah's theorem on the measure problem and related results*(Raisonnier, 1984), whose abstract specifically claims to "give a simpler and metamathematically free proof of Shelah's result"

will allow me to sketch a list of putative objections to calling such proofs metamathematical. These putative objections will shed light on the practice of organizing mathematical proofs by their methodology. Following that, I will present a series of theorems and proofs that involve increasingly substantial use of metamathematical methods. At each turn, I will analyze how the proof in question addresses the earlier objections. This sort of dialectics will hopefully shed light on our intuitive taxonomy of proofs by their methodology, as well as on the specific question of whether a proof can be said to make substantial use of metamathematical methods.

4.2 Philosophical motivations: the nature of categorizing proofs by methodology

Ultimately, I think this exercise will raise more questions than answers. Just so that I do not raise these questions in vain, I would like to situate this work in the context of philosophical discourse surrounding proofs in mathematics. As a point of departure, I would like to first highlight the pervasive practice of characterizing proof as algebraic, combinatorial, topological, etc. That is, characterizing proofs by their methods.

This practice is well-known in the mathematical community and has indeed played a significant role in mathematical progress. For example, G. H. Hardy was well-known to have said that an elementary proof of the prime number theorem (in particular one that does not use methods from complex analysis) would be groundbreaking[3]. In some cases, the quest for an alternative proofs is the main driving force of publications. To wit, Plunkett's 1956 PNAS article (Plunkett, 1956), appropriately titled *A Topological Proof of a Theorem of Complex Analysis*, begins with a remark that demonstrates exactly this point:

[3]For a rough historical sketch, see Goldfeld (2004).

> In a lecture presented by G. T. Whyburn at the 1955 Summer Institute on Set Theoretic Topology, attention was called to the lack of a purely topological proof of the fact that a mapping generated by an analytic function f of a complex variable can be a local homeomorphism at a point z_0 of the complex plane only if $f'(z_0) \neq 0$. It is the purpose of this note to give such a proof of this fact.

Born out of earlier investigations on the practice of reproving theorems (Dawson Jr., 2006; Rav, 1999), Dawson's 2015 monograph *Why Prove it Again* presents an impressive undertaking that collects, compares, and analyzes instances of theorems having multiple proofs, in fields ranging from elementary geometry to mathematical logic. Dawson lists a number of reasons why mathematicians value alternative proofs. These range from practical to aesthetic; they include (Dawson Jr., 2015, Chapter 2):

1. correcting previous errors
2. eliminating superfluous or controversial hypotheses
3. extending the range of validity
4. making the proof more perspicuous
5. purity of method[4] (concern for methodological propriety)
6. benchmarking (demonstrating the power of a given methodology by employing it to prove theorems in areas where it might seem not to be applicable)
7. aesthetic considerations, such as elegance, simplicity, economy, expressions of individual patterns of thought
8. pioneering a new route

[4] I should note that the present work here will likely be topically relevant to the works on purity that are gaining attention in the philosophy of mathematical practice literature, see e.g., Arana (2009), Arana (2017), and Detlefsen and Arana (2011). For the most part, the present project will be orthogonal to the subject of purity, but their interplay will be brought up occasionally.

Elsewhere, in the textbook *Proof and the Art of Mathematics*, Hamkins appeals to the epistemic virtues [5]:

> Nevertheless, we often find it valuable to have multiple proofs of a theorem, especially in the case of an important or central result or in a case where the proofs are extremely different. ... [T]he main reason we value multiple proofs of a theorem is that different arguments, especially when they are extremely different and highlight different fundamental aspects of a topic, deepen our mathematical understanding and appreciation of a mathematical phenomenon. ... Having a proof of a theorem is ... about explaining why the theorem is true, about giving us the conceptual framework in which to understand the mathematical fact more deeply ... reveal[ing] the hidden structure of mathematical reality ... elucidating deep connections between mathematical phenomena ... suggest[ing] different avenues of generalization. (Hamkins, 2020, p. 9)

Of course, as a prerequisite to appreciating the virtues of alternative proofs, one needs an understanding of when two proofs are distinct. Dawson himself sidelines the (difficult) question of when two *formal* proofs are distinct, because the concern here is actual practices of mathematicians, who very seldom consider fully formalized proofs. Like Dawson, here we are interested in this background question of when two proofs, broadly construed, are considered different. We further narrow it down to the question of when can two proofs be considered to have used the same method, or rather, what is it about a method that makes one proof distinct from another that makes use of a different method.

Dawson suggests that "it is usually easy to tell, on informal grounds, whether such proofs are essentially different or merely variants of one another. Intuitively, they are different if

[5] Note that the same book opens with a note to instructors that discourages organizing mathematics "around methods of proof." In that context, "methods of proof" refer to general proof strategies, such as proofs by contradiction or proofs by contrapositives, which has a different meaning than the "methods" we are discussing here or in Dawson's monograph.

they employ different concepts or tactics." Going one step further, I take it that there is also an intuitive consensus of when a proof makes use of some method. For example, a proof that uses the intermediate value theorem uses methods in analysis, and a proof that appeals the pigeonhole principle uses combinatorial methods. Let us stipulate that a proof uses a method from a field of mathematics, if it uses theorems, concepts, or tools originating from that field. The issue of distinguishing proofs by methodology is also of concern to Dawson, who finds fault in Loomis's compendium of proofs of the Pythagorean theorem (Loomis, 1968) in that

> distinctions among proofs are often not clearly or carefully made, despite [Loomis's] proclaimed intent to classify and arrange the proofs according to "method of proof and type of figure used." Indeed, the very first of his 'algebraic' proofs — those based on similarity relations, as opposed to 'geometric' proofs based on area comparisons — appears to differ little ... from 'geometric' proof 230. (Dawson Jr., 2015, p. 25)

However, it is clear that some proofs might ostensibly use different methods, but they are ultimately considered identical ("really the same proofs"). Gowers, writing about the identity of proofs, sketches four proofs, each using a different method, of the irrationality of $\sqrt{2}$: the classical proof by contradiction, another one with infinite descent, one using continued fractions, and finally another one that uses analysis-style considerations of rates of convergence. Gowers goes on to point out why they are really just the same proofs, despite apparent differences:

> For instance, the construction of the sequence (p_n, q_n) in the fourth proof is the same as the construction of the continued-fraction expansion of $\sqrt{2}$ ... Also, the way that we produced (p_n, q_n) from (p_{n-1}, q_{n-1}) is just the reverse of the way that we produced a smaller fraction from $\frac{p}{q}$ in the second proof. The fourth proof is

perhaps very slightly different, in that it involved inequalities, but that was not a fundamental difference. (Gowers, 2007)

For our purpose, Gowers's point is this: two proofs can use different methods, and yet this difference does not necessarily provide us with anything in ways of new perspective or greater generalizability. This motivates yet another stipulation. Say that a proof *makes substantial use* of a method, if it uses the method in a way that contributes to some of those virtues considered by Dawson and Hamkins. From this stipulation it should be clear that if two proofs make substantial uses of different methods, then they can be considered sufficiently distinct, partly because they reveal to us different strains of mathematical depth, to borrow a phrase from Maddy (cf. Maddy (2011)). For example, Furstenberg's celebrated proof of the infinitude of primes makes substantial use of topological methods, marking the beginning of his productive career of applying ergodic-theoretic/topological dynamic methods to number theory. Wiles's proof of Fermat's Last Theorem, of no less fame, makes substantial use of the algebraic methods of Taniyama and Shimura. Gödel's consistency proof of the axiom of choice and the generalized continuum hypothesis makes substantial use of the ordinals and definability- and model-theoretic methods. The French analysts' founding of real analysis made substantial use of the nascent set-theoretic methods. And the investigation of the (co-)analytic sets by Luzin and colleagues makes substantial use of the peculiar methods of sieves and constituents, which were later realized to be the structural tool of well-founded trees and their ranks[6]. Finally, the discovery of pathological sets at the level of Σ^1_2 and Π^1_1 due to Gödel, Novikov, and others was only made possible[7] by metamathematical considerations of Gödel's constructible universe. Such examples abound in the mathematical literature and each one has been a celebrated event that drives subsequent development of entire fields.

To accentuate this point yet again with a soundbite, here is Paul Halmos unreservedly

[6]"Thus did well-founded relations enter mathematical praxis." -Kanamori (1995)

[7]At the end of Luzin (1930), Luzin had declared this problem to be unsolvable, because the concepts involved were logically unrelated.

embracing such practices:

> The existence of eigenvalues is a deep fact, derived by techniques far from the spirit of linear algebra. What it comes down to is that an eigenvalue, a geometric concept, is the same as a zero of the characteristic polynomial, an algebraic concept, and the existence of such zeroes is guaranteed by the fundamental theorem of algebra, an analytic tool. (Halmos, 1978)

Dawson concludes his monograph by calling for further investigation into the multiple-proof phenomena as well as more examples for case studies. The present chapter may be viewed as a partial answer to Dawson's call, working towards more humble sub-questions. For concreteness, we will consider metamathematical proofs in descriptive set theory, especially in cases where a classical proof is available. But rather than pondering what each proof tells us, we will first be concerned with asking whether they are genuinely different proofs and why. And secondly, if, by virtue of making substantial use of metamathematical methods, metamathematical proofs really are its own class of proofs, parallel to algebraic and topological proofs, then what is it that these methods import in a proof, so that they contribute to the foregoing collection of virtues?

Metamathematical methods in descriptive set theory, in my view, present a great opportunity for a case study, because it is not as ingrained on the mathematical community as, say, algebraic methods. It is often intuitively clear when a proof makes substantial use of algebraic methods, as well as what these methods do in a proof. So a similar project involving algebraic proofs will either have little payoff or have to surmount a great number of technical or expository difficulties, given the vast number of algebraic proofs available. Meanwhile, despite there being a handful of metamathematical proofs, it is not so intuitively clear what metamathematical methods import in a proof, even though there exist a handful of proofs using them, with some degree of conceptual unity. This makes them suitable for a conceptual analysis.

On top of that, there seems to be a rather nebulous unease with such methods[8]. This can be seen, for example, in Fenstad and Normann (1974), where Fenstad and Normann specifically reserved a spot for the following remark: "The notion of an absolute $\mathbf{\Delta}^1_2$ set may be too 'metamathematical' for the taste of an analyst. It would be interesting to get some alternative definition of this class." By surveying and analyzing metamathematical methods in descriptive set theory, we hope to dispel some of this unease, demonstrating perhaps the versatility and naturalness of these methods in a family of contexts.

Having situated this project in the broader philosophical context, we now turn to the specific case study of metamathematical proofs. More concretely, we keep three probe questions in mind: one, is it reasonable to assert that there are genuinely proofs of theorems not ostensibly having to do with metamathematics, which are nevertheless characterized by their use of tools, concepts, and techniques originating in metamathematics? (And antecedently, what criteria would license such assertions in other cases?) Two, what is the nature of such uses? In other words, what does a methodology import in a proof? Three, given that a number of these proofs use forcing and transitive models as a way to prove theorems rather than consistency results, what does this say about the equivalence of various ways of understanding forcing, are they still equivalent in this context?

[8] In some cases even a conscious effort to eliminate them, for example Raisonnier's conspicuously titled paper (Raisonnier, 1984) mentioned at the start of this chapter. Or Suppes's sketch of Tarski as "concerened to eliminate ... metamathematical methods in favor of purely mathematical or set-theoretical ones." (Suppes, 1988)

4.3 Metamathematical methods: a preliminary (non-)example and some putative objections

Let me begin with an example that obviously does not make substantial use of metamathematical methods. The purpose of introducing such an obvious non-example is that it will help us come up with a list of putative objections as to why fails to do so. This will facilitate our discussion of what kinds of conditions a proof should satisfy, for us to say that it makes subtantial use of metamathematical methods. For rhetorical purposes, I will perhaps be a little unfair in my scrutiny. The point is that once these objections have been laid out, we will see more clearly how the better examples overcome them in the next section (and how this non-example can redeem itself).

Consider the following folklore proof of the Kleene-Post theorem.

Theorem 4.1. *There are incomparable Turing degrees.*

Proof. (The joke proof). First notice that total comparability of the Turing degrees implies the Continuum Hypothesis: there are continuum-many Turing degrees, and Turing-reducibility places them into an uncountable linear order with countable proper initial segments. This implies that there are ω_1-many degrees and hence ω_1-many reals.

Now $\neg\mathsf{CH}$ can be forced over any model of ZFC, and since "there exists two real numbers, neither of which reduces to the other" is a projective statement of complexity Σ^1_1, by Shoenfield absoluteness implies that it holds to begin with. □

Now, this is a cute argument that apparently use the familiar metamathematical methods, but it gives the impression it makes use of superfluous or unserious use of the tools involved. In fact, one of the very few places this proof is actually written down[9] had to preface it with

[9] Noam Greenberg's personal website (Greenberg, n.d.), at the time of writing

the caveat that it “is really intended as a joke.” At any rate, it gives a similar impression as the following well-circulated “joke” proof of the theorem that $\sqrt[n]{2}$ is irrational for $n \geq 3$: because otherwise $\sqrt[n]{2} = \frac{p}{q}$ would then imply $q^n + q^n = p^n$, contradicting Fermat's Last Theorem!

It is common to describe proofs like these as “nuking a mosquito,” and mathematicians seem to have a good sense of telling when a mosquito has been nuked. It nevertheless benefits us here to reflect on what exactly is funny or unsatisfactory about this proof that makes it unserious, or in our parlanace, fail to make substantial use of metamathematical methods. Hence what follows will be a putative list of the complaints one might launch about it.

Impure methods. There are many things to say about proofs that use high-powered tools to prove simple statements. The most obvious, I think, is that it uses impure methods irrelevant to the problem at hand. This notion here is roughly in line with the purity of methods literature. As Arana & Detlefsen put it succinctly,

> we take the epistemic virtue represented by pure problem solutions - or, more particularly, what we will call topically pure solutions - to be their special potential for relieving specific ignorance. ... The topically determining commitments of a given problem are those which together determine what its content is for a given investigator. ... We say that a solution ... is topically pure when it draws only on such commitments as topically determine [the problem]. (Detlefsen & Arana, 2011)

In light of this, tools like forcing, the Continuum Hypothesis, and Mostowski absoluteness, one might say, are specific to classical concerns of set theory of the reals and cardinality, and so they definitely draw on commitments that are far beyond the concrete grasps of Turing machines. Their applicability, then, would seem almost accidental. Someone with a view like this would find indeed find this proof a bit cheeky: it appears to lead the reader down the carefree candy cane lane of Turing machines, but makes the sudden turn round the corner and out leaps the higher set-thereotic bulldozer, totally obliterating the problem[10].

At this point, I should remind the reader that this chapter was written with the express purpose to study a particular impure method: metamathematical methods in proving theorems distant from metamathematics. By the choice of this subject, topical purity has been a lost cause from the get-go (we will return to this point in later considerations of *dispensibility via translation*). So maybe considerations of purity and relieving specific ignorance will explain the nebulous unease with metamathematics, observed at the end of the previous section. But I do not think it does enough justice to our reflections on the joke proof.

After all, here we fully embrace and celebrate the fact that theorems can be proven by methods that are not topically pure. Consider again Furstenberg's proof of the infinitude of primes using topological methods. This proof is not topically pure, but it is a beautiful proof that has been celebrated for its elegance and fruitfulness. More importantly, and Arana & Detlefsen also acknowledged this in their discussion of it, Furstenberg's proof makes staggering contributions to a majority of the virtues outlined by Dawson and Hamkins. Arana & Detlefsen suggest (but ultimately disagree with) a Bourbakiste attitude, which they ascribe to McLarty, citing private correspondence: "McLarty's view is thus that, properly understood, the topic of [the infinitude of primes] makes room for topological elements, and that it is therefore wrong to classify Furstenberg's proof as impure simply because it makes use of such elements." Although here we are openly not preoccupied with whether the metamathematical methods used are pure or relevant or sanctioned by the problem's commitments, I think McLarty's stance hints at something more relevant to us: there is something off about the joke proof, not because it is impure, but because something else is amiss that prevents it from making substantial use of metamathematical methods. I will try to articulate this next.

[10]One is reminded of a quote by the great comedian Jerry Seinfeld: "Remember writing proofs? That's stand-up. You say something ridiculous and prove it to be true!"

Unsuggestive. Recall one of virtues that Hamkins listed: suggest[ing] different avenues of generalization. The two joke proofs above, especially the use of Fermat's Last Theorem, carry a tone of finality to it[11]. A lowly, innocent problem is introduced, and high-and-mighty tools are brought to bear on it, and the problem is solved. End of story. The proofs do not immediately suggest further research or generalizations, and it fails to do this because, to temporarily use a vague term, it does not contextualize the theorem in a broader framework.

To see an example that does this, consider one of the most successful alternative proofs in set theory: Wadge's proof of the strictly increasing Borel complexities the following sets:

$$K_1 := \{f \in {}^\omega\omega \mid \exists n f(n) = 0\}$$
$$K_2 := \{f \in {}^\omega\omega \mid \forall m \exists n > m\ f(n) = 1\}$$
$$K_3 := \{f \in {}^\omega\omega \mid \exists k \forall m \exists n > m\ f(n) = k\}$$
$$K_4 := \{f \in {}^\omega\omega \mid \forall j \exists k > j \forall m \exists n > m\ f(n) = k\}$$

That these sets have strictly increasing Borel complexities was originally proven by Keldysh in the 1930s[12]. Her results marked "the first significant advance in the study of the Borel classification since the first non-trivial example of a third-class set was given by Baire in 1905" (Chernavskii, 2005).

Let us begin with some simple observations. First, K_1 is open and dense, and so it cannot be closed, lest it be the whole space. Next, K_2 is G_δ, dense, and co-dense, and so by Baire category theorem it cannot be F_σ. Already in the progression from K_1 to K_2, one needs to move from elementary topological considerations to more advanced Baire category tools. In Wadge's own words: "Borel and Keldych gave topological proofs that ... were rather complicated ... and did not seem to generalize to higher levels." (Wadge, 1983, p. 9)

[11]However, we will see that the forcing proof has some redeeming qualities in the business of being suggestive nonetheless.

[12]cf. Luzin (1930, pp. 96–104) for a discussion of K_3, K_4

Compare this with Wadge's genuinely innovative strategy: first, borrowing from Addison's perspective, the topology of Baire space is interpreted computability-theoretically: open sets correspond to recursively enumerable sets, whose membership can be confirmed in finite amount of time; this perspective then allows one to understand the continuous functions on the reals as exact analogues to the computable functions on the natural numbers. Next, taking advantage of infinite games, which were becoming popular at the time, Wadge casts the existence of continuous reduction in terms of having winning strategies for one of the players in a certain 2-player game, the Wadge game as we know it today.

According to Wadge[13], the methods suggested themselves naturally, when Addison supplied him with the computability-theoretic understanding of openness and continuity on the Baire space, as well as a recent publication by Rogers (Rogers, 1959), who obtained the almost entirely analogous result that the following sets have strictly increasing Turing degrees (where $\langle W_i \mid i \in \omega \rangle$ is some effective enumeration of the r.e. sets):

$$R_1 := \{i \in \omega \mid \exists n(n \in W_i)\}$$

$$R_2 := \{i \in \omega \mid \forall m \exists n > m(n \in W_i)\}$$

$$R_3 := \{i \in \omega \mid \exists k(i \in W_k) \land \forall m \exists n > m(n \in W_k)\}$$

$$R_4 := \{i \in \omega \mid \forall j \exists k > j(i \in W_k) \land \forall m \exists n > m(n \in W_k)\}$$

For Wadge, the crucial breakthrough lies not only in the fact that the analogous result can be proven in computability theory, but also that Rogers approached his problems from a different angle, by considering the Turing degrees and many-one reductions. This key insight separates Wadge's proof from Keldysh's in a fundamental way: Borel complexity was now being viewed through the lens of continuous reductions rather than the more topological/algebraic perspective of Luzin and Keldysh: "Of course, there are in the literature many

[13]See Wadge (2020) for Wadge's own account of how he came up with these results.

instances in which continuous preimage is used to derive a particular result. ... Yet nowhere (to our knowledge) is the relation $A = f^{-1}(B)$ ever explicitly defined and studied as a partial order." (Wadge, 1983, p. 3)

Indeed, Keldysh, Borel, Baire, and Luzin were working with improverished means. For example, the Baire space was not considered as a topological space on its own, but rather as proxies for irrational numbers in $(0, 1)$ via continued fractions. To make matters worse, thinking about algebraic operations in terms of logical quantifiers and connectives would not have been available until a few more years later, neither would the very foundations of computability theory. At any rate, for Keldysh, the very motivation of these results was to understand the structure of the Borel sets, which at the time sat squarely at the intersection between analysis and topology. Wadge's proof, on the other hand, was a direct result of the computability- and degree-theoretic thinking, assisted with the formalisms of infinite games. With his alternative proof, Wadge not only greatly simplified upon his predecessors like Luzin and Keldysh, but he also practically placed the problem in a drastically different context from theirs. This immediately suggest numerous natural generalizations and further problems. For example,

> Very soon after the game characterization was discovered, I realized that AD implied the SLO principle. The SLO principle in turn settles the analog of Post's problem. It therefore occurred to Addison and myself that the [Wagde degrees], unlike their recursion-theoretic analogs, might (assuming AD) have a very regular structure. The discovery of the close connection between [Wadge degrees] and the Hausdorff difference hierarchy confirmed this opinion, and so I began the systematic investigation of [the Wadge degrees]. (Wadge (1983))

It is hard to exaggerate the significance of Wadge's new proofs of Keldysh's results. Today, Wadge degrees and Wadge games are some of the most important tools and subjects for descriptive set theorists. The point is clear: one can very sensibly say that Wadge's proof makes substantial use of computability-theoretic and game-theoretic methods, in the sense that presently concerns us[14].

As it stands, the joke proof of the Kleene-Post theorem lacks certain principled insight as to what one can do after reading this proof. In contrast, the classical proof with finite extension methods allows for great degrees of control in the construction and provides a clearer understanding of the two Turing-incomparable sets. Its many rich adaptations would give rise to the later successor of priority methods in computability theory. The metamathematical proof, on the other hand, is trick-y. It is not apparent what work the Continuum Hypothesis is doing, and how we may adapt it to the family of theorems about Turing degrees (We will see later why this is not entirely accurate).

Unexplanatory. A closely related complaint with the joke proof is that it is not quite explanatory. Now, the literature on explanations in mathematics is vast and deep[15], and I cannot do justice to the subject in a few paragraphs here. Nevertheless, I think the reader can agree that the proof seems rather mysterious. It seems like the structure of Turing degrees and the Continuum Hypothesis just happen to be related by a cardinality coincidence. Admittedly, part of the mystery of the joke proof above is adduced by the fact that it does not attempt to demonstrate *why* the theorem is true.

The issue of explanation in descriptive set theory was recently taken up in Antos and

[14]We note for the record that Dawson Jr. (2015) provides an excellent discussion of another famous example from logic, the compactness theorems, in which different proofs of the compactness theorem of first-order logic are considered, starting from Gödel's original proof all the way to the algebraic proofs. Dawson, working towards a similar analysis to the above, writes of Gödel's original proof "...[it is] not perspicuous, because the details of the constructions are rather intricate; and ... depends on forming conjunctions of a finite number of formulas ... and so does not extend to a non-denumerable sequence of formulas." And he proceeds to analyze how later proofs, especially later algebraic ones, bring the familiar algebraic tools to bear, such as ultrafilters and Boolean algebras. These methods freed the problems from the perhaps more stringent confines of syntactic manipulation and semantic valuation, which then revealed intricate connections in areas previously considered distant, natually suggesting a host of open problems.

[15]See Mancosu et al. (2023) for a more comprehensive overview

Colyvan (n.d.). Taking a class of dichotomy theorems[16] as a case study, Antos and Colyvan analyzed two types of proofs: the classical proofs that rely on Cantor-Bendixson-like derivative methods, and the advanced logic proofs that use a combination of forcing and effective descriptive set theoretic arguments. Of particular interest to us is Antos and Colyvan's account of the latter: because of their use of methods apparently unrelated to the dichotomy theorems, such proofs reveal a fruitful source of explanation as to why the dichotomy theorems hold. For example, Antos and Colyvan cited Moschovakis's remark that the perfect set theorem of Σ^1_1 sets (note that this is the lightface version) is susceptible to a recursion-theoretic proof, which provides an explanation in terms of definability rather than cardinality: "an analytic set P has a perfect subset if it has at least one member which is more difficult to define than P itself."

For our present context, this aspect of explanatoriness identified by Antos and Colyvan is closely related to the issue of suggestiveness above. The use of recursion-theoretic techniques immediately suggests the celebrated analogies[17] between classical descriptive set theory and effective descriptive set theory. Namely, such methods "[link] the dichotomy theorems to [effective descriptive set theory] as well as providing a further example for the connection of [classical descriptive set theory] and [effective descriptive set theory]. Thereby it both uses and strengthens the interconnection between CDST and its effective counterpart." Paraphrasing Colyvan et al. (2018), Antos and Colyvan summarized the explanatory values of the advanced logic proof type as follows: "a class of theorems is explained by deriving its members using a proof type that shows that the class of theorems is part of a very general, perhaps utterly pervasive, pattern that is characteristic for the area of mathematics it is a part of."

Part of the discomfort with the joke proof, then, can be understood via this account

[16]For example, Silver's dichotomy theorem which states that a Π^1_1 equivalence relation on a Polish space either has countably many equivalence classes or a perfect set of pairwise inequivalent elements.

[17]See e.g., the Introduction chapter of Moschovakis (2009).

of explanatoriness. We see that the lack of suggestiveness above bleeds into a lack of explanatoriness, ultimately resulting in an unsatisfactory proof. The key defect, in Antos and Colyvan's framework, is that the use of metamathematical elements fails to situate the theorem (or its proof) in a more general pattern characteristic of the study of, say, Turing degrees. We will later see examples that arguably succeed in doing so.

Dispensable via translation. Although it is not obvious at all from the presentation above, the joke proof of the Kleene-Post theorem can be translated, and perhaps sharpened somewhat, to dispense with the use of metamathematical methods. Someone unnerved by talks of generic extensions or transitive models or absoluteness would, for instance, seek to rephrase generic extensions in terms of the syntactic forcing relation[18] and to reduce the final move from consistency to truth by literally going through the normal form of Σ^1_1 relations and showing that (assuming the conclusion does not hold) a particular tree is forced to be ill-founded and well-founded at the same time.

While this kind of translation has little to no payoff in terms of the ideas involved, it does soothe the nerves of those who are uncomfortable with the metamathematical methods, and it helps to ensure that the proof is still technically rigorous, depsite reference to things like forcing extensions and generic filters. This issue appears harmless enough, so we will pause on it until we return to it later, where it becomes more involved with the questions surrounding equivalent ways of understanding forcing.

On the other hand, one might also attempt to translate away the metmathemtatical involvement by really contemplating on what the joke proof is doing. In fact, this is precisely the content of Exercise VII.G8 of Kunen's classic set theory textbook, leaving the following remark preceding it:

> In recursion theory, many classical results may be viewed, in hindsight, as forc-

[18] Incidentally, this is one of the "interpretative strategies" for Barton's Universist in Barton (2020).

> ing arguments. Consider, for example, the Kleene-Post theorem that there are incomparable Turing degrees. ... Furthermore, to conclude recursively incomparability of [the two Cohen reals] ... it is sufficient that [the generic filter] intersect only a few of the arithmetically defined dense sets ... so few that in fact [the two reals] may be taken to be recursive in $0'$.This forcing argument for producing incomparable degrees below $0'$ is in fact precisely the original Kleene-Post argument, with a slight change in notation. (Kunen, 1980, p. 236)

It should be pointed out that Kunen here is explicitly instructing the readers to use the poset adding two mutually generic Cohen reals. This means he is hinting towards a sharper proof than our joke proof above. One needs to make the additional conceptual jump and realize that one of the most canonical ways to force $\neg\mathsf{CH}$ is to add Cohen reals. This sort of translation seems to materially improve on the proof, and in fact in the next section we will see that translation in this specific case actually serving a useful purpose in suggesting further avenues for generalizations.

We will return to the topic of translation after we see more examples. But for now, let me make a temporary point: the metamathematical elements in the joke proof can be reduced by translation. In the present example, doing so will change the proof so drastically that it may be said to be a different proof. However, this issue of dispensability via translation will return to haunt a number of subsequent examples, where it will be tempting to think that these other, better metamathematical proofs can be translated into *equivalent* proofs using only the tools of the relevant field of mathematics. I should warn the reader that this is one of the more serious challenges to the assertion that those proofs make substantial use of metamathematical methods.

To drive the point home: the joke proofs above make use of high-and-lofty power tools without adding much beyond the existing classical proofs. This means a few things. One, conceptual complexity is increased without much gain in the direction of informativeness or generalizations; and two, the heavy machinery is used, which can be dispensed with by a translation into lighter tools without loss. In the next section, I shall present better examples that will hopefully address these objections.

4.4 Some counterpoints and better examples

4.4.1 A more serious proof and its relatives

As I have confessed at the beginning, I have exaggerated a little with the defects of the joke proof for rhetorical purposes. Although on face value it does not immediately suggest further generalizations, it does lead to interesting directions that can potentially contribute to the list of virtues in the beginning of the chapter. For example, if one takes the equivalence between the Continuum Hypothesis and the total comparability of the Turing degrees seriously, the use of forcing will suggest that it may be fruitful to study the computability-theoretic properties of the generic sets added to violate CH. Guided by this heuristic, let us consider, following Kunen's exercise, a second attempt to use metamathematical methods to prove the Kleene-Post theorem.

Proof. (The more serious proof). Force to add two mutually generic Cohen reals c, d. Obviously c, d are Turing-incomparable, because otherwise (say) $c \leq_T d$ would imply that $c \in V[d]$, contradicting mutual genericity. And so the extension has incomparable Turing degrees. Again, this is absolute to models of set theory, and so it holds to begin with. □

In my view, the more serious proof is more suggestive than the joke proof. For one thing, it explicitly produces two Turing-comparable sets (albeit in the generic extension) and hints at what kinds of objects will be useful in thinking about Turing-incomparability. From this proof, it is also immediately clear how to obtain an anti-chain in the Turing degrees, even in the presence of stronger notions of computability. This is done, for example, in Theorem 6.2 of Hamkins and Lewis (2000): having previously established that the assertion "there exists countably many reals that are pairwise incomparable under infinite-time-Turing-machine reducibility" has complexity Σ^1_2, Hamkins and Lewis immediately derives it by adding ω many mutually generic Cohen reals, concluding that the assertion is true in the generic extension and hence true in the ground model by absoluteness.

Moreover, having the concrete forcing poset in hand, one also gets a rough idea as to why there are Turing-incomparable sets. Someone reading the more serious proof will understand that mutually generic Cohen reals are in some sense computationally weak. They cannot compute each other, for instance. Now this very naturally suggests the question of what they can compute, leading to the following lemma.

Lemma 4.2. *Let M be a countable transitive model of enough of ZFC, and let x be a real in M and c a Cohen real over M. If x is computable relative to c, then x is computable.*

Blass. If x is computed by the Turing program Φ^c_e, then this fact also holds true in $M[c]$, and so by the forcing theorem this is forced by some condition p. That is,

$$p \Vdash \text{ the } \check{e}\text{th Turing program in the oracle } \dot{c} \text{ computes } \check{x}$$

For any $i \in \omega$ we compute $x(i)$ as follows: run $\Phi^s_e(i)$ for all the s extending p.

As soon as any of these computations halt, the output will be the correct value of $x(i)$. This is because: if s_0, s_1 are two different nodes extending p and $\Phi^{s_0}_e(i) = 0 \neq 1 = \Phi^{s_1}_e(i)$, then we can build two different filters G_0 and G_1 containing s_0, s_1 respectively. Now $M[G_0]$ and $M[G_1]$ will both think x is computed by Φ^a_e (since both filters contain p. Note that they will interpret a differently; but that doesn't matter). So $M[G_0]$ thinks that $x(i) = 0$ and $M[G_1]$

thinks $x(i) = 1$. But whatever $x(i)$ is, this is an absolute fact about $x \in M$, so it should be answered in the same way by all transitive models extending M. Contradiction! □

Considering Solovay's characterization that the set of Cohen reals over a countable transitive model is comeager, one obtains the next corollary in a straightforward manner.

Corollary 4.3 (Hinman, Thomason, Feferman)**.** *If a real is computable relative to a comeager set of reals (i.e., its Turing cone is comeager), then it is computable.*

Proof. Let x be a real whose Turing cone is comeager. Since the comeager sets form a filter, every comeager set must contain Cohen reals (over some countable transitive model), and so x is computable from a Cohen real. By the previous lemma, it is computable. □

The previous examples are meant to demonstrate why the more serious proof is more fruitful than the joke proof. The joke proof, while amusing, does not immediately suggest any further generalizations or directions of inquiry. Contemplating it a little bit more, however, as Kunen suggested, we arrive at the more serious proof, which has a better claim to have made substantial use of metamathematical methods.

Adding to the evidence, we turn to the well-known descriptive-set-theoretic theme duality between measure and category in descriptive set theory (cf. Oxtoby (1980)). Suppose, again, that one takes seriously the idea that forcing to violate CH will lead to interesting computability-theoretic objects. Then, knowing that this has led to the conclusion that Cohen reals are computationally useless and knowing the measure-category duality, one arrives straightforwardly at the next lemma.

Lemma 4.4. *Let M be a countable transitive model of enough of ZFC, and let x be a real in M and c a random real over M. If x is computable relative to c, then x is computable.*

From which the same argument as before, relying on the weak homogeneity of random forcing instead of Cohen forcing, gives the following corollary.

Corollary 4.5. *If a real is computable relative to a measure one set of reals (i.e., its Turing cone has measure 1), then it is computable.*

Following the suggestions of the more serious metamathematical proof, we have thus arrived at a proof of an important theorem in computability theory and descriptive set theory. The corollary as stated was first discovered by de Leeuw, Moore, Shannon, and Shapiro (Leeuw et al. (1956)) and remains a celebrated result about the robustness of the Church-Turing Thesis (see remarks after Theorem 8.12.1 in Downey and Hirschfeldt (2010)). And its hyperarithmetic analogue (i.e., a real whose hyperarithmetic cone has full Lebesgue measure is hyperarithmetic), proved by Sacks, is a cornerstone of the theory of hyperarithmetic sets (cf. chapters II and IV of Sacks (2016)).

Hopefully, the preceding discussion demonstrates that the more serious proof is more suggestive compared to the joke proof, and hence can be characterized as making more substantial use of metamathematical methods than the latter. I should point out that, on the explanatory front, the more serious proof also situates the theorem in the broader projects of understanding the measure-category duality (Oxtoby, 1980) on the one hand, and the computational/definitional strength of random/generic reals on the other. Just as the use of effective techniques in the proof of perfect set theorem for analytic sets explains why analytic sets have perfect subsets in terms of how difficult it is to define its elements, the use of metamathematical tools in this case illustrates the pervasive pattern that generic and random objects are computationally weak, and hence do not compute each other (cf. Chong and Yu (2015), Chong et al. (2008), and Monin and D'Auriac (2019)).

4.4.2 Mycielski's perfect set theorem

The intricate connections with computability and with measure and category we just witnessed above merit a deeper look. In the previous series of examples, the metamathematical tools of forcing, transitive models, and absoluteness provide a versatile set of suitable tools to study definability and the measure-category duality. This is not an isolated occurrence, but instead a recurring theme in descriptive set theory. For example, consider the following theorem of Mycielski, with a metamathematical proof.

Theorem 4.6. *Let $R \subseteq X^2$ be a Borel equivalence relation on a Polish space X, such that each equivalence class is meager. Then there exists a perfect set of pairwise inequivalent elements.*

Proof. Force to add a perfect set of mutually generic "Cohen reals". In any extension, the interpretation of the Borel code of R is still an equivalence relation with meager equivalence classes. This is because being an equivalence relation is a Π^1_1 property, and the equivalence classes being meager is equivalent to the relation itself being a meager subset of X^2 (this follows from Kuratowski-Ulam), which is a Σ^1_2 property.

Now consider an arbitrary Cohen real c in that perfect set. Already in the intermediate extension $V[c]$, c belongs to an equivalence class that is meager, which is in turn contained in an F_σ meager set by the usual properties of Baire category. Since this F_σ meager set is coded in $V[c]$, any Cohen real over $V[c]$ will not be in that set, by Solovay's characterization of Cohen-genericity. This includes all Cohen reals on that perfect tree added by the forcing. Therefore, any two such Cohen reals are R-inequivalent, and there is a perfect set of them.

Finally, the statement that there is a perfect tree, any two branches of which are R-inequivalent, is Σ^1_2 in the code of R and hence absolute to V. □

As with the last few examples, this proof makes substantial use of tools such as Solovay's characterization of genericity, generic extensions, complexity calculations, and absoluteness. Moreover, the proof of Theorem 4.6 sees a higher degree of involvement of the metamathematical tools. First, a new cast member, the Borel codes, plays a role in the proof that did not exist in our previous examples. Second, the proof engages in more serious complexity calculations: for instance the property of being meager, or the statement of the theorem's conclusion. Furthermore, these tools engage more heavily with the concepts involved in the theorem, as seen by the use of Kuratowski-Ulam theorem to obtain the Σ^1_2 complexity, as well as appealing to the F_σ-covering property of meager sets to show that the Cohen reals are R-inequivalent.

Noticing that the theorem in fact entails the existence of Turing-incomparable reals (in fact, a perfect set of them) by simply taking R to be Turing-equivalence, we see that it inherits much of the same explanatory power as the previous examples. As before, the proof roughly illustrates the recurring theme that mutually generic objects carry no information about each other, as far as the equivalence relation R is concerned, and so what previously prevented them from computing each other now prevents them from reaching each other via R, so to speak.

Incidentally, the proofs we have looked at so far all share a similar template: to prove a statement P, one establishes that P (or another statement that implies it) is consistent by methods of forcing and/or inner models; the complexity of P is then shown to be simple enough to not be affected by such methods, and finally one uses absoluteness to conclude that P holds. For conciseness, let me call this the *Field-of-Dreams* template: if you build it (i.e., the requisite model), he (i.e., the theorem) will come (true). The next two examples use the absoluteness theorem as an intermediate step, rather than the finishing move.

4.4.3 No Borel reduction from eventual equality to identity

Notice that Mycielski's Theorem 4.6 effectively produces a continuous injection from the Cantor space into the Polish space X, with the additional property that distinct reals are mapped to R-inequivalent elements. This fact can be understood in terms of Borel reducibility.

Definition 4.7. *Let E be a Borel equivalence relation on a standard Borel space X and similarly F a Borel equivalence relation on a standard Borel space Y. We say E is Borel reducible to F (written as $E \leq_B F$) iff there is a Borel function $f : X \to Y$ such that $x_1 E x_2 \Leftrightarrow f(x_1) F f(x_2)$. Such a function f is called a Borel reduction of E to F.*

In other words, Theorem 4.6 shows that the identity relation on Cantor space is Borel reducible to the relation R on the Polish space X. The next theorem shows that the converse is not true in general.

Theorem 4.8. *There is no Borel function $F : 2^\omega \to 2^\omega$ such that $xE_0y \Leftrightarrow F(x) = F(y)$, where E_0 is the equivalence relation of being different in only finitely many places.*

Proof. Suppose towards a contradiction that F is such a function, and let b_F be its Borel code. Now force to add a Cohen real c. In $V[c]$, the function F^* coded by b_F still has the same properties as in the assumption of the theorem, by Π^1_1-absoluteness. But now in $V[c]$, the image $w = F^*(c)$ of the Cohen real under this map remains the same regardless finite changes to c, which implies the value of w is already decided by the weakest condition (suppose not, then pick two incomparable conditions s, t of equal length that force disagreeing values to this image. For any Cohen real extending s, the same tail extending t is another Cohen real. But these two Cohen reals will be mapped to different images, contradicting that F maps finitely-different reals to the same real).

So w is already in the ground model, and so its pre-image $F^{-1}(w)$ will contain a real that differs from a Cohen real in only finitely many places. But this is impossible, as the Cohen real is generic over the ground model. □

4.4.4 Friedman's Borel diagonalization theorem

In H. Friedman, 1981, Friedman gave a forcing proof of the following theorem.

Theorem 4.9. *There is no uniform Borel diagonalizer. That is, there is no Borel function* $F : \mathbb{R}^\omega \to \mathbb{R}$*, such that for all* $g \in \mathbb{R}^\omega$*, and all* $n \in \omega$*, we have* $F(f) \neq f(n)$*; and that if* $\operatorname{ran} f = \operatorname{ran} g$*, then* $F(f) = F(g)$*.*

In words: Cantor's diagonalization cannot be performed in a Borel way that respects permutations of the given sequece. Or in slightly imprecise words, there's no Borel way to diagonalize out of any given countable *set* of reals (because $\operatorname{ran}(f) = \operatorname{ran}(g)$ means f and g enumerate the same set).

Proof. Suppose towards a contradiction that there is such a Borel map F. Forcing with $\operatorname{Col}(\omega, \mathbb{R})$ to make the ground model reals countable, let f and g be mutually generic. In $V[f][g]$, the re-interpreted map F^* still satisfies the assumption by absoluteness. But since f and g enumerate the same set of reals (i.e., the ground model reals), we have that $F^*(f) = F^*(g)$, which implies that $z := F^*(f) = F^*(g)$ belongs to both $V[f]$ and $V[g]$. By Solovay's lemma on intersection of extensions from mutual generics, we obtain that $z \in V$, which is a contradiction since F is suppose to diagonalize out of the ground model reals. □

Friedman remarks that "However in the Appendix we give a proof using the Baire category theorem applied to $\underset{\sim}{R}^{\mathbb{N}}$, where $\underset{\sim}{R}$ is the reals with discrete topology. The two proofs are essentially equivalent." In this case we even seem to have a reversal of fortune: the forcing proof is sharp and informative, situated naturally in the context of the problem. The Baire category proof uses rather concocted topology, with no straightforward suggestions in sight.

4.5 Interlude: philosophical reflections on the examples

Having seen a number of examples, it would appear that metamathematical proofs contribute to many of the virtues of having multiple proofs, in ways that the more familiar examples do. Can these proofs be said to make substantial use of metamathematical methods? We now pause to make good on a few debts accrued earlier the project. We shall be presently occupied with two strands of questions. One, are we, like in the relative consistency proofs, free to choose from any of the equivalent ways understand the use of forcing in these proofs? And two, can the metamathematical elements be translated away without loss of content? These will be addressed in turn.

4.5.1 How to prove something true by proving it consistent

Proofs of the *Field-of-Dreams* type are a curious class of arguments, distinguished by their use of forcing as a way to prove something to be true rather than consistent. This is made possible by well-known absoluteness results like Shoenfield absoluteness, which entail that statements of certain complexities are absolute between models of set theory, and in particular between a model and its generic extensions.

Despite their talk of generic extensions, proofs like these can be made rigorous in the usual ways. Experts familiar with how this is done can feel free to skip the following and jump to the end of this subsection. Kunen (2013), for example, lists two ways[19] in which

[19]Curiously, Kunen brushes over how this can be done in the Boolean-valued approach to forcing. This can be done, for example, by using the kind of Boolean ultrapower considered in Hamkins and Seabold (2012). Once one has all the requisite $j : V \prec \bar{V} \subseteq \bar{V}[G]$, all that's left to do is to verify that the assertions

this can be done.

The countable transitive model (ctm) method. To prove a statement φ, first take a large enough $H_\theta \prec_{10000} V$ and a suitable countable elementary submodel $M \prec H_\theta$. One then forces over M (or rather, its transitive collapse) to obtain $M[G]$, so that one has literally two models of (enough) set theory to use the absoluteness theorems. Once it is shown that $M[G] \vDash \varphi$, absoluteness implies $M \vDash \varphi$, and so by elementarity φ holds in V.

The syntactic method. Here one dispenses talks of generic extensions altogether by defining a relation $\Vdash^*$ and proving that it satisfies all the properties of the usual forcing relation. Schematically, for each statement φ known to be absolute across models of set theory, one shows that $p \Vdash^* \varphi$ iff $1 \Vdash \varphi$ iff φ.

I am writing this down in part for the chapter to be self-contained. But more importantly, I wish to draw attention to a key philosophical issue. For consistency results, any one of these ways of interpreting talks of generic extensions are just as good as any other. Based on the foregoing paragraphs, the situation is roughly the same with the *Field-of-Dreams* proofs that we have surveyed so far. It would seem it is only a matter of taste how one chooses to formalize them; in particular, the use of countable transitive models can be avoided without loss of generality. Let us keep this in mind as we proceed, as examples in the next section will illustrate that this may not always be the case with metamathematical methods.

4.5.2 Revisiting dispensability by translation

Another lingering question: are the metamathematical tools doing any real work in these proofs? What if the use of the metamthemtaical methods is merely *une façon de parler*?

What I am suggesting is the possibility that the metamathematical methods are in fact not

in question are absolute between $\bar{V} \subseteq \bar{V}[G]$, which may or may not be transitive anymore.

contributing to the mathematical content behind the proofs. That is, one may translate the metamathematical proofs into "purely mathematical" proofs, without loss of their relevant content. We will consider this possibility here. Ultimately it will turn on how we translate the proofs.

We have seen how some of the metamthematical appearances in the proofs can be reduced by a literal translation. This can be done by, say, translating forcing into something more syntactic or algebraic to dispense with talks of generic extensions or models, or interpreting various absoluteness theorems as structural results about the relevant pointclasses, not about immunity of sentences of certain logical complexities under forcing and inner models. The benefit of doing this is mostly just therapeutic for the faint-hearted and someone excessively concerned with formal rigor. Let me call this harmless act of translation a literal translation.

On the other hand, let us return to Kunen's hint for his exercise G8, reproduced below:

> In recursion theory, many classical results may be viewed, in hindsight, as forcing arguments. Consider, for example, the Kleene-Post theorem that there are incomparable Turing degrees. ... Furthermore, to conclude recursively incomparability of [the two Cohen reals] ... it is sufficient that [the generic filter] intersect only a few of the arithmetically defined dense sets ... so few that in fact [the two reals] may be taken to be recursive in $0'$.This forcing argument for producing incomparable degrees below $0'$ is in fact precisely the original Kleene-Post argument, with a slight change in notation. (Kunen, 1980, p. 236)

Upon hearing someone suggest that the more serious forcing proof of the Kleene-Post theorem makes substantial use of metamthematical methods, one might be tempted to object that the forcing proof is essentially *just* the finite extension method, with more bells and whistles wrapped in the language of forcing. Such an objection would then proceed along Kunen's lines, attempting to isolate the pieces of the proof that are really doing the heavy lifting, in this case the two Cohen reals. Following Kunen, one would then recognize that, for the two Cohen reals to be not Turing-comparable, it is sufficient that the generic filter intersect only a very small portion of the dense sets. In fact, once we understand what the dense sets we need to meet, then we will quickly realize that we can outright prove that these two reals exist. At this point the proof is slowly morphing into the more classical proof by finite extension, and the "essentially just the same proofs with different notations" objection seems more damning.

Let me call this more material type of translation, strategic translation. In a strategic translation, the main idea of the proof is identified, and the relevant objects and properties are rephrased in a lean-and-clean way that completely dispenses with the metamathematical tools. It would be particularly challenging to the claim that a proof uses metamathematical methods in a sui generis way if the proof can be translated away in this manner. In particular, if a proof using an alternative method can be strategically translated back to a more classical proof without loss of the mathematical content, then we seem to have good reasons to say that the alternative method has not been used substantially, that the two proofs are essentially the same proof[20].

It is instructive here to also compare the metamathematical proof of Theorem 4.6 with its classical proof.

Sketch of the classical proof of Theorem 4.6. Let R be a Borel equivalence relation on a Polish space X such that each equivalence class is meager. Again, this is equivalent, by Kuratowski-Ulam, to R being a meager subset of X^2. Thus there exists a countable sequence of open dense sets $U_n \subseteq X^2$ such that $R \cap \bigcap_n U_n = \emptyset$.

Fixing a complete compatible metric for X, we define a collection of open sets $(V_s)_{s\in 2^{<\omega}}$

[20] This is reminiscent of the more formal foundational investigation of theory comparison via translation. See for example Meadows, 2023b and Meadows, 2023a for the different grades of theory equivalence by means of translation.

by induction on lh(s), maintaining that the following conditions are satisfied:

(i) $\mathrm{diam}(V_s) \leq 2^{-\mathrm{lh}(s)}$

(ii) for $i \in \{0,1\}$, $\mathrm{cl}(V_{s^\frown i}) \subseteq V_s$

(iii) for all $n \in \omega$, for all $s \neq t \in 2^{<\omega}$ with $\mathrm{lh}(s) = \mathrm{lh}(t) = n+1$, $V_s \times V_t \subseteq \bigcap_{m \leq n} D_m$

And finally the map $x \mapsto \bigcap_{n\in\omega} V_{x\restriction n}$ is a continuous injection from 2^ω into X, whose image is a perfect set of pairwise inequivalent elements under R. □

Kunen's hint on how to translate away the metamathematical elements in the proof of the Kleene-Post theorem lends itself to a similar perspective on how the metamathematical proof is really just the classical proof in a different framework, that the metamathematical elements in the proof are not doing the real work. To see this, notice that the perfect set of Cohen-generic (in the sense of forcing) reals, for example, is replaced with an explicit construction of a perfect set of sufficiently generic (in the sense of Baire category) elements. Nevertheless, the construction stays close to the spirit of forcing, in that it essentially uses Baire-category-like topological arguments to construct the open sets $(V_s)_{s\in 2^{<\omega}}$, the same way that the generic filters are supposed to be constructed (using the Rasiowa-Sikorski lemma). Similarly, the classical proof places a tighter control on what dense sets are supposed to be met, such that the generic elements can be outright proved to exist, exactly as in the case of the finite-extension method used in Kleene-Post theorem's classical proof.

Taking stock: although the preceding examples of proofs using metamathematical methods address the issues of suggestiveness and explanatoriness that the joke proof suffered from, we see that they are falling short of making substantial use of metamathematical methods, because such methods can be dispensed with by a strategic translation, preserving the main idea of the proof. Heeding the issues raised in the brief interlude, I will submit the following series of examples that I think will show some degree of indispensability by strategic translation, as well as the unavoidability of countable transitive models.

4.6 More serious examples

Keep in mind that we are concerned with two questions: one, whether the use of forcing can be understood in any of the equivalent ways, in particular that the use of countable transitive models can be in principle avoided without loss of generality; two, whether the metamathematical elements can be translated away strategically without loss of content; that is, if a proof that uses metamathematical methods is really just a more complicated way of saying something that can be said in a more classical way.

I claim that the examples in this section will tend towards a negative answer on both fronts.

4.6.1 Measurability of choice sets of WO

Theorem 4.10. *Say two reals x, y from WO, the set of reals coding well-orderings, are equivalent, just in case code well-orderings of the same ordertype. Let $A \subseteq 2^\omega$ be such that it contains one and only one element from each equivalence class. Then A is measurable. In fact A has measure zero.*

Metamathematical Proof, Fenstad and Normann (1974). Let M be an arbitrary countable transitive model of (enough of) ZFC. So now $A = W_0 \cup W_1$, where W_0 codes the ordinals in M and W_1 codes those not in M. Now, W_0 is a countable set of reals, and hence has measure zero. Next we show W_1 can be covered by a countable union of measure zero sets, which implies that A has measure zero.

Consider random forcing over M. We claim that any real $r \in W_1$ will be non-random overe M. This is because if it were, then the generic extension $M[r]$ of M would have the same ordinals as M, and hence the ordinal coded by r is in M, contradicting that $r \in W_1$.

Now since each $r \in W_1$ fails to be random over M, by Solovay's characterization of random-genericity, r belongs to a Borel set of measure zero coded in M. But there can be only countably many such sets, so W_1 is covered by a countable union of measure zero sets. $\square$

We start off by noticing that the countable transitive model approach to forcing is used essentially in this proof, in that one appeals to M's countability to ensure the null-ness of W_0. Moreover, generic extension and generic filters are understood literally, so that Solovay's characterization of randomness can be leveraged to provide a null covering of W_1. Interpreting forcing in any of the other equivalent ways will forestall the proof strategy. Also, amusingly, to establish the non-randomness of the reals in W_1, the current proof also appeals to a peculiar feature of forcing: that it does not add ordinals.

These observations together imply that the use of forcing, and especially the ctm approach, cannot be straightforwardly translated away. We include this example here because it marks a significant departure from the *Field-of-Dreams* proof type in this regard. We have previously outlined how a *Field-of-Dreams* proof may be interpreted, without loss of rigor or content, in any of the equivalent approaches to forcing. This is no longer the case with the current proof.

For convenient comparison, the classical proof found in Luzin and Sierpiński (1918) is included. The original proof was written in the language of sieves and constituents. I will present a modernized version of it.

Classical Proof of Theorem 4.10. First notice that WO can be partitioned into Borel set

$\{P_\alpha \mid \alpha < \omega_1\}$, where each P_α is the set of reals coding well-ordering of type α. Second, since WO is Π^1_1, it is measurable, and by usual properties of Lebesgue measure, $\mathsf{WO} = \bigcup_{n\in\omega} N \cup M_n$, where N has measure zero and each M_n is closed.

By Σ^1_1-boundedness, each M_n is bounded in WO. Write α_n as the lease uppoer bound of (the ordinals coded in) M_n. Note that this implies that for all $\beta > \alpha_n$, we have $M_n \cap P_\beta = \emptyset$. In other words, $M_n = \bigcup_{\alpha<\alpha_n} M_n \cap P_\alpha$.

But now observe that, since $P_\alpha \cap A$ only has a single element, $M_n \cap A$ is at most countable and hence measure zero. Therefore, $A = A \cap \mathsf{WO} = \bigcup_{n\in\omega}(A \cap N) \cup (A \cap M_n)$. This writes A as a countable union of measure zero sets, and hence A has measure zero. □

Now, the classical proof and the metamathematical proof in this case still share some similarities in the proof strategies; for example, both proceed by partitioning A into an obviously measure zero set and a set that is ultimately shown to have measure zero as well, so arguably this can be seen as a case strategic translation is still available between the two. With the next proof (folklore), finally, we shall find an example in which the metamathematical methods are indispensable in a more profound way.

4.6.2 Measurability of analytic sets

Theorem 4.11. *If $A \subseteq \mathbb{R}$ is $\underset{\sim}{\Sigma}^1_1$, then A is measurable.*

Proof. Suppose $A := \{x \in \mathbb{R} \mid \varphi(x, a)\}$, where φ is Σ^1_1 and $a \in \mathbb{R}$. Now consider the random forcing poset $\mathcal{B}/\mathrm{Null}$. Let X be a G_δ set that forces $\varphi(\dot{r}, \check{a})$ (or rather, its equivalence class in the random forcing algebra does), where $\dot{r}$ is a canonical name of the random real added. Note that X can be assumed to be G_δ because of general properties of Lebesgue measure.

Claim: $\mu(X \triangle A) = 0$. (This is just the equivalent formulation of the measurability of A.)

To see the claim, assume towards a contradiction that, say, $B = A \setminus X$ has positive outer measure (the case where $X \setminus A$ has positive outer measure is similar). Pick some H_κ large enough, so that it contains a, A, X and reflects the relevant facts, such as that $H_\kappa \vDash [\mathbb{R} \setminus X] \Vdash \neg\varphi(\dot{r}, \check{a})$. Now let $\pi : N \to H_\kappa$ be an elementary embedding where N is countable transitive with $a, X, A, B \in \mathrm{ran}(\pi)$. Write $\pi(\bar{A}) = A, \pi(\bar{X}) = X, \pi(\bar{B}) = B$.

It follows that there is a real $r \in B$ random over N. Notice that $[B]$ is a stronger condition than $[\mathbb{R} \setminus X]$. And we have $N[r] \vDash \varphi(r, a)$, since $r \in A$ by assumption and Σ^1_1 formulas are absolute between V and $N[r]$. But this last fact contradicts that $N \vDash [\bar{B}] \Vdash \neg\varphi(\dot{r}, \check{a})$, because with $r \in B$ we would also have $N[r] \vDash \neg\varphi(r, a)$. $\square$

Before I comment on this particular proof, observe that the metamathematical proofs of Theorems 4.11 and 4.10 both immediately suggest how to prove that the relevant sets have the property of Baire almost as an afterthought: just use Cohen forcing instead of random forcing. Mutatis mutandis, they also show that these sets are *universally measurable*, i.e., μ-measurable with respect to every σ-finite, complete, and regular Borel measure μ. This is in line with the lesson we have learned so far, that metamathematical methods are especially versatile in dealing with the pervasive patterns of duality between measure and category, making them as explanatory as the proofs of dichotomy theorems using effective techniques in the case of Antos and Colyvan (n.d.).

Returning to the proof at hand, the familiar tools of forcing, countable transitive models, and absoluteness are again in play, but they are doing more substantial work than in the examples from the previous sections. This is best appreciated by recalling the usual, more classical proof of analytic measurability[21] that proceeds via a structural analysis of the analytic sets, showing that they are the results of applying the Suslin operation, which preserves measurability, to the closed sets. In contrast, the metamathematical proof relies

[21] See e.g., Theorem 11.18 in Jech (2003).

crucially on the definitional complexity of analytic sets, and applies to absoluteness to this characterization of them. In addition, generic extensions are taken literally, over countable transitive models, whose countability licenses the use of Solovay's characterization of random-genericity to find actual random reals that allow the proof to go through.

The interactions between forcing, countable transitive models, absoluteness, and measurability are on full display in the current example. Unlike the previous examples, it is sufficiently different from the classical proofs, with no straightforward strategic translation in sight, and there is no substitute for the use of countable transitive models. Therefore, it seems reasonable to say in this case we have a proof that genuinely makes substantial use of metamathematical methods.

4.7 Inconclusive discussions and further directions

In this chapter, I considered the use of metamathematical methods in descriptive set theory to prove theorems that are relatively distant from metamathematics. Starting with obvious nonstarters, I moved on to proofs with increasingly serious involvement of metamathematical tools, especially the Field-of-Dreams type that leverage logical complexity and absoluteness. I then questioned if this metamathematical involvement is merely an illusion that can be dispensed by translation. Then I cited Solovay-type characterizations of genericity and claimed that metamathematical methods are a natural framework to study things like the measure-category duality, and that in some cases, metamathematical considerations provide genuinely new mathematical insight. Finally, I appealed to metamathematical proofs that cannot be translated away, taking them to be the final piece of evidence that there is a sui generis methodology to metamathematical proofs. In other words, some metamathematical proofs do contribute geneuinely new information/insights to theorems that can otherwise by obtained by standard methods, and they do so in a way that is almost entirely analogous to other cases of innovative proofs using new methods.

I began the chapter by predicting that it would raise more questions than answers. I would like to consider myself successful in this regard. First, granting that my analysis of characterizing proofs by their methodology is correct, and that metamathematical methods do constitute a sui generis method parallel with other better-known methods, this raises interesting questions about the nature of the tools used. In Barton (2020), on the topic of Universism versus Multiversism in set theory, the fruitful applications of generic extensions in proving theorems (beyond mere relative consistency results) and formulating axioms places additional constraints on the Universist in her interpretations of forcing, if she is to maintain a certain degree of naturalness or transparency in her interpretation:

> [S]ituating V within a Forcing Multiversist framework (i.e. someone who thinks that no universe is maximal and we can always move to a forcing extension) allows us to solve important and difficult set-theoretic questions ... Thus, forcing extensions can function with respect to V in contemporary set theory somewhat like the historical situation with complex and real numbers..

Should the practical utility of generic extensions, inner models, and absoluteness in proving theorems in descriptive set theory be taken to be evidence one way or the other in the more philosophical debates about the philosophy of set theory? This more question looms large beyond the scope of this chapter. I should only point out that we have seen examples of proofs that are only possible when one interprets forcing to be a technique associated with countable transitive models, and that this data point should be taken into account in the broader philosophical debates.

Relatedly, Williams (2022) raised the question of whether bi-intepretability results allow us to say that structuralist set theories, like the Elementary Theory of Category of Sets, and the more material set theories like ZF are really equivalent ways of looking at the same thing.

An important piece of evidence stems from the observation that Shoenfield Absoluteness allows one to show that, e.g., a number-theoretic statement provable with the axiom of choice can be proved without it. So a ZF theorist would be privy to the knowledge that, for example, the use of idempotent ultrafilters in proving Hindman's theorem is unnecessary. An ETCS theorist on the other hand, while also having apparent access to a translated version of Shoenfield Absoluteness via bi-interpretation, would supposedly be disadvantaged. This is because all known proofs of Shoenfield Absoluteness "make essential use of the particularities of material set theory", in particular encoding the ordinals in a Δ_0 fashion, namely the von Neumann ordinals. This makes possible a crucial step in showing that Π^1_1 statements are absolute, by reducing their truth to the well-foundedness of trees, which can now seen to be Δ_1.

Having sketched what we have called a literal translation of a proof of Shoenfield Absoluteness in ETCS, Williams deemed it unsatisfactory, because the proof

> doesn't avoid the particularities of material set theory - a global membership relation, the Foundation axiom, the von Neumann encoding of ordinals, etc. These play an essential role in the proof. The reason ETCS + Replacement can carry out this proof is precisely because it can internally mimic material set theory.

From this, Williams suggests that "the use of material set theory isn't mere bookkeeping convenience to make proofs shorter and easier. Rather, it provides essential mathematical content."

I take it that our case study shares much of Williams's concerns. This ranges from the broader issue of whether two technically equivalent proofs can be said to carry the same mathematical content, to the more specific issue of the indispensability for particular elements involved in the proofs (von Neuman ordinals, countable transitive models, etc.)

In this chapter we have seen stronger examples in close spirit to Williams's, where the additional mathematical content extends beyond just what axioms are/are not necessary to prove a theorem, but rather new methods of proving theorems altogther. How this might inform Williams's challenge to the structuralist is beyond the scope of this chapter and deferred to future work.

On a different note, much of what was said and analyzed can also be applied to the case of nonstandard methods, using nonstandard models of arithmetic or nonstandard analysis to prove theorems of interest to "standard" arithmetic and analysis. The last two decades or so saw a proliferation of attention to such methods[22], and yet their efficacy still remains in debate, or at least suffers from misimpressions. Much of the acrimony hinges precisely on whether nonstandard methods are just roundabout ways of using standard methods after all, that it is dispensable via translation. For example, Katz and Katz (2010) reports that Paul Halmos, having seen an early reprint of the Bernstein-Robinson proof of the invariant subspace conjecture, rushed to produce a translation in standard techniques and published it in the same issue of the journal as the Bernstein-Robinson proof. He would go on and describe the distinction between nonstandard methods and standard methods "a matter of personal preference, not necessity" (Halmos, 1978).

In fact, this impression was so widespread that Henson and Keisler's influential 1986 had to start with the following proclamation: "It is often asserted in the literature that any theore be proved using nonstandard analysis can also be proved without it ... and thus there is no need for nonstandard analysis", as well as the explicit purpose "to show that this assertion is wrong."

Here we see the same issues of dispensability via translation being echoed as in our putative objections to the metamathematical methods. It will be worthwhile and practically relevant for future work to investigate whether the same kind of analysis can be applied

[22]For recent contributions, see e.g., Di Nasso et al., 2015, 2019; Duanmu et al., 2021

to nonstandard methods, or maybe also to the broader recent success of model-theoretic methods in other areas of mathematics, and see what kinds of conclusions can be drawn[23].

[23] Regrettably, the author realized in hindsight that nonstandard methods probably would have been a better target of this project. Alas, it will have to be deferred to another work.

Chapter 5

Pre-history of Invariant Descriptive Set Theory

5.1 Introduction

Descriptive set theory is the deep and robust mathematical theory dealing with definability in Polish spaces. With its intellectual ancestry in the works of French analysts Borel, Lebesgue, and Baire, the theory was born, under curious circumstances, out of Suslin's correction of a simple mistake in Lebesgue's *Sur les fonctions représentables analytiquement* (Lebesgue, 1905, pp. 191–192). From 1910 to the 1930s, it emerged as a distinct field, being then a subfield of topology that studies the Borel sets, analytic sets, and projective sets in general[1]. Since the 1950s, intricate connections between descriptive set theory, recursion theory, and the theory of inner models[2] and large cardinals made it a salient tool in mathematical logic,

[1] According to Koepke (n.d.), the name "descriptive point set theory" ("deskriptive Punktmengenlehre") first appeared in Alexandroff and Hopf (1935), where it was defined as the study of these particular sets: "Zur Topologie gehört auch die *deskriptive Punktmengenlehre* (Théorie descriptive des ensembles), d.h. im wesentlichen die Theorie der Borelschen Mengen, der A-Mengen und der projektiven Mengen."

[2] For an excellent discussion of why and how inner model theory needs descriptive set theory, see the MathOverflow discussion Atmai (2015)

especially the study of the axiom of determinacy.

The foregoing paragraph forms the basis of what I shall call the common story of descriptive set theory, usually found in existing literature such as Ferreirós (2023) and Kanamori (1995), the introduction chapter to the classic textbook Moschovakis (2009), in philosophical discourse such as Maddy (1988b, 2000), as well as the historical survey Moschovakis (2018) (except for a one-sentence ending of the latter that reads "starting with Harrington-Kechris-Louveau (1990), the most important results in DST have been applications to many areas of mathematics").

Fascinating as it may very well be, this story tends to sideline a range of developments that led to an outpouring of descriptive set theory publications in the last three decades, significantly elevating its presence in the general mathematical community today. Indeed, a cursory glance at some of the recent publications in generalist mathematics journal such as *Annals of Mathematics* and *Inventiones Mathematicae* reveals that the field has been making remarkable contributions to a wide range of mathematical areas, including ergodic theory (Foreman et al., 2011), functional analysis (Sabok, 2016), measure theory (Conley & Miller, 2017), and graph combinatorics (Bernshteyn, 2023; Marks & Unger, 2017).

The subject matters in these papers, and more importantly the tools and methods employed in dealing with them, are somewhat distant from the original focus of descriptive set theory. In particular, one will find very little mention of forcing, large cardinals, or inner models which occupy much of the set-theoretic discourse today, and it gets little motivation in ways of classical concerns such as the continuum hypothesis or axiom justification. And yet they occupy a respectable position in the descriptive-set-theoretic landscape today, suggesting that the common story of descriptive set theory is to be supplemented with a more complete picture.

I am of course referring the rich theory, developed in the past 30 years, of invariant descriptive set theory[3]. Being the subject of Kechris's programmatic 1998 Gödel Lecture (Kechris, 1999), invariant descriptive set theory is descriptive set theory's venture into the study of definable equivalence relations on Polish spaces, typically focusing on the structures and complexity among various classification problems and establishing various reducibility and non-reducitbility results. As Kanovei puts it in Kanovei (2008, p.ix), "classification problems for different types of mathematical structures have been in the center of interests in descriptive set theory during the last 15-20 years."

By all accounts, invariant descriptive set theory looks like a miraculous sudden appearance. The abstract framework of Borel equivalence relations began with the two publications H. Friedman and Stanley (1989) and Harrington et al. (1990), seemingly out of nowhere. With the framework in place, previously disparate theorems across various fields could then be unified and re-interpreted as structural comparisons among the difficulty of classification problems. Part of the motivation of the present chapter is to try to dispel this air of mystery surrounding the birth of invariant descriptive set theory, and to show that many fundamental results in the field can find some sort of spiritual ancestor in the early developments of descriptive set theory.

Now, the mathematical contexts immediately preceding the birth of the theory of Borel equivalence relations are well-known to experts in the field and are scattered in the introduction sections in various specialist literature. This places certain restrictions on where original contributions can be made. So let me narrow down my humble goal: to identify precursors to the theory of Borel equivalence relations, perhaps pre-dating the materials commonly found in the relevant specialist literature.

Inevitably, parts of the chapter will involve presenting various technical facts already

[3]A better known name of the subject is Borel equivalence relation theory (or the theory of definable equivalence relations), such as in the title of Hjorth (2010) in the Handbook of Set Theory. *Invariant descriptive set theory*, a term (re-)gaining popularity these days, is the title of the sole textbook Gao (2008) on the subject to date. I opted for this latter term based on its precedence (J. Burgess & Miller, 1975; D. E. Miller, 1976) and because it is more concise, although I will often use the two terms interchangeably.

known to experts. Hopefully this is forgivable, since this is not intended to be a mathematical chapter. Its value, and the goal, is to trace the pre-history of these facts and to play folklore archaeologist, if you will.

A few words on motivation are in order. First, like I mentioned above, invariant descriptive set theory marked such a distinct thread in the development of set theory that it naturally warrants a closer look of how it came about. Meanwhile, related historical claims abound in the technical literature. For instance, Kechris in his lecture notes (Kechris, 2024) remarked that "the subject has a long history in the context of ergodic theory and operator algebras". Elsewhere, Gao claimed (Gao, 2008, p. 117) that the Borel reduction concept was "originally borrowed from computability theory and structural complexity theory to help determine the relative complexity of equivalence relations". For a statement of more historical significance, Kanovei (1985) claimed that Luzin was " at the source of this trend [to study equivalence relations] ... was probably the first in the descriptive theory to turn his attention to the difficulties associated with equivalence relations." While these claims are more likely true than not, it would nevertheless be beneficial to conduct an historical investigation corroborating them, which is what I hope to do in this chapter.

At the same time, the history of descriptive set theory, especially of the development of different proofs of the so-called *dichotomy theorems* (a family of theorems closely related to the study of definable equivalence relations), has figured in recent philosophical works on explanation in mathematics. For instance, the recent paper Antos and Colyvan (n.d.) compares the putative explanatory powers of the proofs of such theorems, in reference to existing accounts of explanatory values[4]. One main task of Antos and Colyvan (n.d.) is to place the earlier proofs (the so-called advanced logic proofs) and the more recent ones in the broader historical context in which they were discovered. To do this, part of the paper surveyed the history of the dichotomy theorems and the related development of descriptive

[4]See Chapter 4 for a closer look at this work and how it informs the philosophy of mathematics.

set theory, following Miller's historical exposition in B. D. Miller (2009, 2012), while noting piecemeal the sources of each proof and their historical background. Another motivation for this chapter, then, is to provide a systematic and readily available resource, in case the need for such a historical account arises in future investigations.

5.2 Pre-history of invariant descriptive set theory: what is known

For the sake of completeness and being self-contained, I will begin by reviewing what the current literature has to offer. In doing so, I will be brief with the details already well-documented in the literature and refer the interested reader to existing surveys[5]. Readers familiar with the literature may wish to skip parts of this section as they see fit.

One of the earliest motivations for descriptive set theory can be traced back to the works of Cantor, whose *Beitrag* introduced the first formulation the Continuum Hypothesis (CH): every set of real numbers can either be put in bijection with the natural numbers, or with the real numbers. Immediately after, Cantor alluded to an inductive procedure by which the CH can be proved:

> Via an inductive procedure (*Inductionsverfahren*), whose presentation we shall not detail here, the proposition is suggested that the number of classes resulting from this principle of division[6] is finite, specifically, that it is equal to two [...]

[5]There are a number of excellent surveys on this history. For an excellent overview, see Kanamori (1995). A few articles focus on the development of descriptive set theory under the works of particular mathematicians, e.g., Hausdorff (Koepke, n.d.), Luzin (Kanovei, 1985; Keldysh, 1974; Uspenskii, 1985), or groups of mathematicians, e.g., the Moscow school (Novikov & Lyapunov, 1948). Also see Maddy (1988b) and Kanovei (1988) for a survey of the modern developments in relation to determinacy hypotheses and large cardinals.

[6]It is curious to note that the original formulation of the continuum hypothesis was in terms of equivalence classes, namely if one partitions the sets of real numbers in terms of having one-to-one correspondence, then

We will postpone a thorough examination of this question to a later occasion. (Cantor, 1878, p. 257)

The prospects of such an inductive procedure lent itself to a "concrete approach to his continuum problem" (Kanamori, 1995, p. 245). It is often remarked[7] today that the inductive procedure Cantor had in mind referred to some kind of induction on the complexity of point sets, and for each complexity class, proving that no counterexample to the CH can come from that class, perhaps by means of the later perfect set property[8]. This is sometimes informally referred to as Cantor's program, for example in Hamkins (n.d.) and Schindler (2011).

For its principled approach to inductively establish the CH along certain measure of complexity, Cantor's program marked the first instance of a mathematical development guided by a particular complexity-regularity principle that would underpin much of the development of descriptive set theory in the coming decades. This principle is what Moschovakis calls the common motivation of descriptive set theory (who roughly attributes it to Lebesgue): "Constructively defined sets and functions should have special properties that distinguish them from arbitrary ones" (Moschovakis, 2016a). In other words, sets of simple complexity should enjoy certain kinds of regularity properties. For Cantor's program, the requisite regularity property is the perfect set property, whereas the relevant measure of complexity would have to wait until the works of the French analysts, notably Borel, Baire, and Lebesgue.

The advent of a liberal concept of functions[9], as well as the nascent set-theoretical meth-

---

the number of equivalence classes is the lowest possible, i.e., two. However, this is likely unrelated to the later attention to Borel equivalence relations

[7]Of course, Cantor never made explicit what his *Inductionsverfahren* was. Referencing a letter from Cantor to Vivanti, Hallett reconstructs the argument as basing on facts of cardinal arithmetic instead (Hallett (1988, pp. 85–86)). Regardless of what Cantor himself had in mind, actual attempts to systematically establish the continuum hypothesis in the following decades fell roughly in line with the characterization in this paragraph.

[8]A set $X \subseteq \mathbb{R}$ has the perfect set property if and only if it is either countable or has a perfect subset, i.e., a (non-empty) closed subset with no isolated point. The relevant implication here is that a set with the perfect set property has the cardinality of the continuum, and thus cannot witness the failure of the CH.

[9]I am referring to the understanding of functions as a special kind of arbitrary pairing of elements. The history of such an understanding is well-known, see for example Kline (1990, Chapter 44)

ods, brought along a number of pathological sets and functions, and late-19th-century and early-20th-century mathematics paid considerable attention to these pathological objects. Counterexamples to naive intuition proliferated, some of which, like Peano's space-filling curve or the Weierstrass function, are now part of standard pedagogy.

With considerable foresight, Borel adopted a converse approach, motivated by his project to develop a general notion of measure. Borel (1898) delineated the collection of the *ensembles mesurables*, in the first appearance of the later Borel sets:

> The sets whose measure can be defined by virtue of the preceding definitions shall be called the measurable sets. [Their] essential properties, summarized below because they shall be useful to us, are as follows: the measure of the sum (i.e., union) of countably many sets is equal to the sum of their measures; the measure of the difference of two sets is equal to the difference of their measures; the measure is never negative; any set whose measure is not zero is not countable ... Moreover, it is expressly understood that we will speak of measure only in connection with the sets that we have called measurable.[10] (Borel, 1898, p. 48)

In trying to tame sets of reals, Borel adopted a context-restriction approach, having gone so far as claiming (Borel, 1898, p. 48) that a definition of measure for other sets would not be so useful, especially if such a definition does not endow the measurable sets with the aforementioned "essential properties". At the same time, Baire (1899) considered on the other hand a stratification of functions: with the continuous functions sitting at level 0, the Baire class 1 functions are pointwise limits of continuous functions, and the Baire

[10]"Les ensembles dont on peut définir la mesure en vertu des définitions précédentes seront dits par nous ensembles mesurables ... Ces propriétś essentielles, que nous résumons ici parce qu'elles nous seront utiles, sont les suivantes; La mesure de la somme d'une infinité dénombrable d'ensembles est égale à lu somme de leurs mesures ; la mesure de la différence de deux ensembles est égale à la différence de leurs mesures; la mesure n'est jamais négative; tout ensemble dont la mesure nést pas nulle nést pas dénombrable ... Il est d'ailleurs expressément entendu que nous ne parlerons de mesure qu'à propos des ensembles que nous avons appelés mesurables." Translations are my own unless otherwise noted.

class 2 functions are pointwise limits of Baire class 1 functions, and so on. More generally, for a countable ordinal α, a function f belongs to Baire class α just in case there are countably many functions $f_0, f_1, f_2, \ldots$, all belonging to some lower Baire classes, such that $f(x) = \lim_{n\to\infty} f_n(x)$.

The systematic correspondence between the two was worked out in Lebesgue's influential memoir (Lebesgue, 1905). In addition to showing that Baire's classes are proper (that is, there exists new functions at each level that do not belong to the lower levels) and peter out at ω_1, Lebesgue identified the Borel sets as pre-images of open sets under Baire's functions, effectively endowing the former with a hierarchical stratification. This was the first instance of a complexity hierarchy in descriptive set theory, foreshadowing the early moniker "descriptive theory of functions"[11].

With the Borel sets having been shown to satisfy Lebesgue measurability and the property of Baire, a number of questions suggested themselves. Against the backdrop of controversies surrounding the axiom of choice, Lebesgue summarized a number of constructive results as "naming" (*nommer*) particular sets or functions. For example, he summarized his method of universal sets to prove the existence of new functions on every level of Baire's hierarchy as "we can name a function that is neither in class α nor in any lower class" (Lebesgue, 1905, p. 208), and he closed the memoir by posing the famous question of whether one can name a non-measurable set, having previously named a measurable non-Borel set by giving it an analytic expression[12].

Aside from applications of the newfound technology of analysis, measure, and category,

[11] This was used interchangeably with the term "descriptive set theory" for a while, notably in the writings of Luzin and Kuratowski. Hausdorff and Brieskorn (2008, p. 10) attributes the first occurrence of the term "descriptif" for mathematical content to Charles Jean de la Vallée Poussin: "...I address the questions of a more exclusively descriptive nature ... of the distribution of functions in the successive classes of Baire, his theorems on the functions of class 1, and the extensions of that theorem owing to Lebesgue." (de la Vallée Poussin, 1916)

[12] "Donc, nous savons nommer un ensemble mesurable qui n'est pas mesurable B. Mais la question beaucoup plus intéressante: peut-on nommer un ensemble non mesurable? reste entière." (Lebesgue, 1905, p. 216)

two thematic threads emerged out of the works of the French analysts. The first was the problem of determining the range of the sets satisfying the regularity properties, or rather finding pathological sets that can be explicitly constructed. The second: to practice mathematics as constructively as possible, operatively diminishing the use of the axiom of choice[13]. As it shall become clear, the former played a significant role in the subsequent development of Luzin's research program, and the latter a primary motivation for Sierpiński's efforts to bring the axiom of choice into the mathematical canon. Both of these consequences will be shown to play a non-trivial role in the birth of invariant descriptive set theory.

The French analysts' work on the Borel and Baire hierarchy in the abstract was subsequently taken up by the Moscow school of functions, notably under the leadership of Luzin, whose contributions to mathematics, and more specifically to descriptive set theory, are well-documented. See for instance Kanovei (1985), Keldysh (1974), Phillips (1978), and Uspenskii (1985). Relevant to us is his systematic development of the theory of the projective sets, motivated in part by Suslin and other's discovery of the analytic sets and the successful demonstration of their regularity properties.

A veritable tour de force, the comprehensive Luzin (1930) summarized the works of the Moscow school of functions on the theory of the projective sets, setting the tone for later research by proposing, for instance, the measurability problem of the projective sets, the uniformization problems, and the separation principles, etc. Luzin concluded his 1930 with brief philosophical reflections on the profound difficulties associated with such problems, before issuing a rousing call to arms to pursue the mathematics, cementing the theme of the regularity-complexity interplay in the project of descriptive set theory:

> Ultimately, the question [the structure and regularity properties of the projective sets] will only be definitively resolved through the efforts of scientific thought, that

[13]This is a bit of an oversimplification. The constructive attitude here is sometimes called *semi-intuitionist* in the literature and is more philosophically involved. See for instance the relevant sections of Troelstra's survey (Troelstra, 2011)

is, through the observation of mathematical facts. Philosophical considerations are always vague and only serve to distinguish a truly fruitful direction from an infinity of others. Only two cases are possible.

Either subsequent research will one day lead to precise relationships between projective sets as well as to the complete solution of questions relating to the measure, category, power of these sets. From this moment, projective sets will have gained citizenship in Mathematics, in the same way as the most classic Borel sets.

Or the problems indicated on the projective sets will remain forever without solution, augmented by a number of new problems as natural and as unaffordable. In this case it is clear that the day would have come when we shall reform our ideas about the arithmetic continuum.[14] (Luzin, 1930, p. 324)

At the same time, Kuratowski and Tarski (1931) observed that the operations defining the projective sets correspond systematically to logical operations. In particular, the geometric operation of projection could be viewed as the logical operation of existential quantification, a groundbreaking conceptual move that justified the name descriptive set theory.

Research activities in descriptive set theory ground to a slow halt in the Eastern Bloc, partly due to insurmountable metamathematical difficulties (independence from ZFC) and partly due to the infamous Luzin Affair[15]. In the west, Gödel was able to observe that his constructible universe provides a solution to Luzin's measure problem. Namely, there exists

[14]"En fin de compte, la question ne sera résolue définitivement que par les efforts de la pensée scientifique, c'est-à-dire par l'observation des faits mathématiques. Les considérations philosophiques sont toujours vagues et ne servent qu'à distinguer une direction vraiment fructueuse d'une infinité d'autres. Deux cas seulement sont possibles. Ou bien les recherches ultérieures conduiront un jour aux relations précises entre les ensembles projectifs ainsi qu'à la solution complète des questions relatives à la mesure, catégorie, puissance de ces ensembles. A partir de ce moment, les ensembles projectifs auront conquis droit de cité en Mathématiques, au même titre que les plus classiques des ensembles mesurables B. Ou bien les problèmes indiqués sur les ensembles projectifs resteront à jamais sans solution augmentés de quantité de problèmes nouveaux aussi naturels et aussi inabordables. Dans ce cas il est clair que le jour serait venu de réformer nos idées sur le continu arithmétique."

[15]The interested reader is referred to Demidov et al. (2016) for a detailed account of the history.

a Σ^1_2 set in L that is not Lebesgue measurable. This interplay between inner models and regularity properties was to be the foundation of the modern theory of determinacy and inner models of large cardinals. Much of this part of the history is well-documented, I refer the reader to Kanamori (2003).

Independently, outside of mathematical logic, a number of researchers engaged in projects of classification of various types. For example, von Neumann (1932), studying the behavior of measure-preserving diffeomorphisms, deferred a question for future work. The program set forth by von Neumann, according to Foreman (2020)[16], is to "Classify the Lebesgue measure preserving diffeomorphisms of compact manifolds up to measure preserving isomorphism."

Several celebrated results in ergodic are the upshots of this program. The Kolmogorov-Sinai entropy, for example, was shown by Ornstein (1970) to be a complete invariant for their isomorphism problem: there is a computable procedure assigning real numbers to Bernoulli shifts, such that two shifts are isomorphic if and only if the two real numbers associated with them are equal. Decades earlier, Halmos and Von Neumann (1942), the sequel to von Neumann's 1932, had achieved a similar result for the discrete spectrum measure preserving transformations, associating with them countable sets of complex numbers that serve as complete invariants for the isomorphism problem of the former[17].

Elsewhere, developments in the study of group actions led Glimm (1961) and later Effros (1965) to prove their famous dichotomy theorem, with heavy measure-theoretic and Baire-category machinery:

Theorem 5.1 (The original Glimm-Effros Dichotomy, see Corollary 6.2.4 in Gao (2008))**.** *Let G be a Polish group that acts on the Polish X. Suppose the orbit equivalence relation E^X_G is F_σ. Then either E^X_G can be completely classified by identity on the real numbers, or else it can serve as the complete invariants for the E_0, the equivalence relation of eventual*

[16]To which I refer the reader for technical background and a survey of related results.
[17]See Hjorth (2000) for a more comprehensive survey of the technical details

agreement on 2^ω.

The famed topological Vaught's conjecture placed additional focus on the study of Polish group actions and their orbit equivalence relations in particular, which states that for any Polish group G acting on a Polish space X, either there are countably many orbits, or there are perfectly many orbits.

With the outpouring results concerning the classification of various types of mathematical structures, as well as their intimate connections with Polish group actions, the stage was set for the birth of invariant descriptive set theory[18]. To wit, H. Friedman and Stanley (1989), citing Vaught's conjecture, and independently Harrington et al. (1990), citing the dichotomy theorem by Glimm and Effros, introduced the abstract study of definable equivalence relations on Polish spaces. With the introduction of the abstract notion of Borel reduction (recall Definition 4.7), previous results where able to be viewed under the lens of complexity comparison among equivalence relations, and general structural results concerning this complexity space could now be established.

5.3 Traces in early descriptive set theory

Having briefly sketched the events leading to the birth of invariant descriptive set theory, we notice that are two distinct threads undergirding its development. On the one hand, the interplay between complexity and regularity, engendered in the works of Luzin and his Moscow school, and on the other, the study of classification problems stemming from e.g., ergodic theory and operator algebras.

[18] I should clarify that the choice of terminology here is a bit anachronistic. The term "invariant descriptive set theory" was in use prior to the 1990s, in association with works on Vaught's conjecture. See the introduction to Becker and Kechris (1996) for a brief overview of the history. At any rate, here we use the term interchangeably with "the theory of Borel (and other definable) equivalence relations"

For the most part, these two threads seem unrelated. And so in this section we address the question: prior to Harrington et al. (1990) and H. Friedman and Stanley (1989), did descriptive set theory pay attention to equivalence relations? And if so, did it do so in any manner foreshadowing the practices of modern invariant descriptive set theory, i.e., with attention to definability and reducibility? We see two main sources of such early traces: Sierpiński's program and Luzin (1927).

5.3.1 Sierpiński's program and the first result in Borel equivalence relations

To find such traces, we need to go back to early days of set theory, especially controversies surrounding the axiom of choice.

Earlier debates about the axiom of choice tended to proceed on philosophical grounds, as evidenced in the *Cinq Lettres* (Hadamard, 1905). However, in the 1910s, Sierpiński initiated a series of investigations into the role of the axiom of choice in mathematics. Rather than justifying the axiom of choice on philosophical grounds like his predecessors, Sierpiński's goal was to identify the use of the axiom of choice in as many theorems as possible, and in doing so, to show the pervasiveness of the axiom in parts of mathematics that are already accepted. Perhaps pre-dating Gödel's later identification of extrinsic justifications of axioms, Sierpiński's wrote in his 1918:

> 1. numerous particular cases of [the axiom of choice] have verified themselves (these cases can be verified independently); 2. one has drawn from the axiom of M. Zermelo a crowd of consequences of which none have led to a contradiction; 3. the axiom of M. Zermelo is indispensable for the demonstration of a great number of important theorems of set theory and of analysis; ... Thus whether we

> are personally disposed to accept the axiom of M. Zermelo or not, we must take account of its role in set theory and in analysis.[19]

Sierpiśki's' goals are most clearly articulated in Sierpiński (1965, p. 100):

> It is most desirable to distinguish between theorems which can be proved without the aid of the axiom of choice and those which we are not able to prove without the aid of this axiom. Analysing proofs based on the axiom of choice we can 1. ascertain that the proof in question makes use of a certain particular case of the axiom of choice, 2. determine the particular case of the axiom of choice which is sufficient for the proof of the theorem in question, and the case which is necessary for the proof, ... 3. determine that particular case of the axiom of choice which is both necessary and sufficient for the proof of the theorem in question.

As part of this effort, Sierpiński (1917) sets out to show that intuitively obvious ("bien démontré") principles used in real analysis already imply the existence of non-measurable sets. Among the principles studied is the following.

Theorem 5.2. *If the set of countable subsets of the reals has the cardinality as the continuum ($|[\mathbb{R}]^\omega| = |\mathbb{R}|$), then there is a non-measurable set.*

Sierpiński proved this by noticing that the assumption of an injection $f : [\mathbb{R}]^\omega \to \mathbb{R}$ would map distinct Vitali classes to distinct reals, a map that cannot be measurable:

> Now let x be a given real number. Let $E(x)$ denote the set of all numbers $x + r$, where r is any rational number. It is easy to see that this will be a countable set, and that we will always have $E(x) = E(x')$ for $x - x'$ rational and $E(x) \neq E(x')$ for $x - x'$ irrational.

[19] Translation from Therrien (2020)

To every given real number x, there exists a corresponding real number $\varphi(x) = f[E(x)]$, and it follows from the properties of $E(x)$ and $f(E)$ that we will have $\varphi(x) = \varphi(x')$ for $x - x'$ rational and $\varphi(x) \neq \varphi$ for $x - x'$ irrational.

I claim that any function $\varphi(x)$ enjoying this property is non-measurable[20]. (Sierpiński, 1917, p. 882)

Let me briefly sketch Sierpiński's argument.

Lemma 5.3. *Any map $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\varphi(x) = \varphi(y)$ iff $x - y \in \mathbb{Q}$ is non-measurable.*

Proof sketch. Assume $\varphi : \mathbb{R} \to \mathbb{R}$ is measurable and $\varphi(x) = \varphi(y)$ iff $x - y \in \mathbb{Q}$. Then $\psi : \mathbb{R} \to \mathbb{R}$ defined by $\psi(x) = \varphi(x) - \varphi(-x)$ is also measurable. Now the two sets $A := \{x \mid \psi(x) > 0\}, B := \{x \mid \psi(x) < 0\}$ are both measurable. And $\psi(x) = 0$ iff $x \in \mathbb{Q}$. So A and B are symmetrical about every rational point ($\psi(2r - x) = -\psi(2r + x)$). And they have the same measure in every open intervals with rational endpoints. And so A and B each occupies half of each rational interval, contradicting Lebesgue density. □

Recalling Definition 4.7, a map $\varphi : \mathbb{R} \to \mathbb{R}$ satisfying the condition $\varphi(x) = \varphi(y)$ iff $x - y \in \mathbb{Q}$ is exactly a reduction from the Vitali equivalence relation to Identity on $\mathbb{R}$. Hence already in 1917, Sierpiński had announced, in modern parlance, that there is no measurable reduction from Vitali equivalence to Identity[21]. Note also that he attributes an earlier, weaker result to Lebesgue, citing Lebesgue (1907), which directly addressed Zermelo's proof of the well-ordering theorem.

[20]"Soit maintenant x un nombre réel donné. Désignons par $E(x)$ l'ensemble de tous les nombres $x + r$, r étant un nombre rationnel quelconque: on voit sans peine que ce sera un ensemble dénombrable et que nous aurons toujours $E(x) = E(x')$ pour $x - x'$ rationnel et $E(x) \neq E(x')$ pour $x - x'$ irrationnel. A tout nombre réel donné x correspondra donc un nombre réel $\varphi(x) = f[E(x)]$, et il suit des propriétés de $E(x)$ et $f(E)$ que nous aurons $\varphi(x) = \varphi(x')$ pour $x - x'$ rationnel et $\varphi(x) \neq \varphi$ pour $x - x'$ irrationnel. Or, je dis que toute fonction $\varphi(x)$ jouissant de cette propriété est non mesurable."

[21]A general result entailing reduction in the other direction would later be discovered by Mycielski (1964). This is Mycielski's Theorem 4.6.

Observe that a fortiori this implies that there can be no Borel reduction. An amusing proof of a variant of this latter fact, making use of metamathematical methods has already appeared in Theorem 4.8 Chapter 4. We adapt the proof here: suppose to the contrary that there is a Borel reduction F. To say that it is a Borel reduction is to say a Π^1_1 sentence about its Borel code c_F, so by Mostowski absoluteness c_F still codes such a reduction in a forcing extension adjoining a Cohen real r. Since any rational translate of r is still Cohen-generic, the image $w = F(r)$ (of the re-interpreted map in the extension) does not depend on the choice of generics, and so the image already exists in the ground model. But now this implies that a ground model real is a rational distance away from a Cohen real, which is impossible.

So as early as the 1960's, we have already seen traces of results about the complexity of classification problems, in particular that Vitali equivalence is strictly more complex than Identity. Another way to canonically increase the complexity of equivalence relations is via the Friedman-Stanley jump, a construction whose clues can be traced back to Sierpiński (1954), where the following theorem is given:

Theorem 5.4. *The following are equivalent:*

1. $\aleph_1 \leq |\mathbb{R}|$
2. *There is a diagonalizer for countable sets of reals: a function* $F : [\mathbb{R}]^\omega \to \mathbb{R}$ *such that* $F(X) \notin X$.

Proof. For the forward direction, assume $i : \aleph_1 \to \mathbb{R}$ is injective. Then given a countable subset of the real numbers X, let $F(X)$ be $i(\alpha)$ for the least α such that $i(\alpha) \notin X$. This is a diagonalizer.

Conversely, if F is a diagaonalizer, we recursively define an ω_1 sequence of reals r_α:

$$r_0 = F(\emptyset)$$

$$r_\alpha = F(\{r_\beta : \beta < \alpha\}) \cup \{r_\beta : \beta < \alpha\}$$

Such a sequence is injective and witnesses $\aleph_1 \leq |\mathbb{R}|$. □

With $\aleph_1 \leq |\mathbb{R}|$ already known to be a choice principle, Sierpiński was able to show that so too is the existence of a diagonalizer for countable sets of reals must involve the axiom of choice. It would have well been within the spirit of Sierpiński's program to conjecture that such a map cannot be concrete. This is in fact the content of Friedman's Borel diagonalization theorem, a prototypical instance to the later Friedman-Stanley jump.

Theorem 5.5 (See also Theorem 4.9)**.** *Define the equivalence relation $\sim$ on $\mathbb{R}^\omega$: $S \sim T$ iff $rng(S) = rng(T)$. Then there is no Borel map $F : \mathbb{R}^\omega \to \mathbb{R}$ satisfying*

1. $S \sim T \Rightarrow F(S) = F(T)$
2. $\forall S \forall n (F(S) \neq S(n))$

That is, there is no (uniform) Borel diagonalizer.

So, despite not being primarily concerned with a general theory of equivalence relations, Sierpiński program to bring the axiom of choice into mathematical canon inadvertently demonstrated that equivalence relations can engage fruitfully with concepts like measure and category. And in doing so, he provided some of the earliest results in Borel equivalence relations.

For a separate source concerning early traces of definable equivalence relations, we now turn to Luzin's work.

5.3.2 Luzin (1927) and the "Common Motivation" again

Whereas Sierpiński was primarily concerned with the use of the axiom of choice in actual mathematics, Luzin proceeded on more principled grounds, following the "Common Motivation" outlined in the introduction of this chapter. In his 1927, the topic of the axiom of choice is reserved for chapter XIV ("Le transfini"), which begins with a discussion of partitions and ways to choose elements from them.

For Luzin, if there is a "precise manner by means of a completely finite law" to obtain a set containing exactly one element from each set in a partition, then the choice is a *Lebesgue choice* ("choix lebesguien"). Otherwise, the choice is Zermelian, meaning that the axiom of choice is invoked.

Luzin singles out the case of a partition with countably many equivalence classes (Luzin, 1927, p. 81), where a Lebesgue choice is indeed imposed ("bien imposé"), since it is not possible to number the classes themselves with the natural numbers without having a precise way to number the points in those sets in some similar way. Of course, with the benefit of hindsight we know this is not true (Luzin's assertion amounts to the axiom of countable choice). The point is that Luzin can be observed to hold partitions with countably many equivalence classes in a privileged position, for similar reasons that he holds the nicely definable sets in a similar position.

This is further corroborated by his later distinction between uncountable and uncountable partitions. In the case of uncountable partitions, Luzin is more circumspect. He writes:

> Thus, there is an essential difference between the case where M [the set of equivalence classes] is numbered by means of positive integers and the case where M is numbered by means of real numbers. The possibility of having the sets M' numbered by means of real numbers without having the Lebesgue choice lies,

> in our opinion, in the fact that the continuum cannot be effectively put into a well-ordered form[22]. (Luzin, 1927, p. 82)

Nevertheless, one special case of uncountable partitions is singled out:

> It is quite different in the singular case where we can draw from R a finite law λ which defines a set of points L satisfying the following two properties: 1. xRx' is false whenever x, x' are distinct points in L; 2. For any point y in the continuum, there exists a point x in L such that xRy is true.
>
> We call a partition with these two properties a *Lebesgue partition*. In this case, the totality T [the set of equivalence classes], being completed, exists in reality; it is therefore legitimate.
>
> But, in the general case where we no longer have the Lebesgue partition, the totality T is, in our opinion, completely illegitimate: it is nothing more than a pure virtuality[23]. (83-84)

Having previously characterized the Borel sets as "legitimate" (in Luzin (1927, pp. 37–38) and later again in Luzin (1930, p. 18)), here Luzin can be understood as describing two types of equivalence relations that are supposed to enjoy a more privileged status, in similar ways that the Borel sets are supposed to be more "legitimate" than the general sets. One is equivalence relations having only countably many equivalence classes; the other is the

[22]"Ainsi, il y a une différence essentielle entre le cas où M est numéroté au moyen des entiers positifs et le cas où M est numéroté au moyen des nombres réels. La possibilité d'avoir les ensembles M' numérotés au moyen des nombres réels sans avoir le choix lebesguien tient, à notre avis, à ce que le continu ne peut pas être mis effectivement sous la forme bien ordonnée."

[23]"Il en est tout autrement pour le cas singulier où nous pouvons tirer de R une loi finie λ qui définit un ensemble de points L ouïssant des deux propriétés suivantes: 1. xRx' est fausse si les points x et x' $(x \neq x')$ appartiennent à L; 2. Quel que soit le point y pris dans le continu, il existe un point x de L tel que xRy est vraie. Nous appellerons partage lebesguien tout partage qui possède ces deux propriétés. C'est dans ce cas seul que la totalité T existe réellement, étant achevée; elle est donc légitime. Mais, dans le cas général où nous n'avons plus du partage lebesguien, la totalité T est, à notre avis, tout illégitime: ce n'eut qu'une pure virtualité"

so-called *Lebesgue partitions*, or in more modern terms: equivalence relations with perfectly many inequivalent classes.

Considering much of descriptive set theory developed under the guidance of the Common Motivation: "Constructively defined sets and functions should have special properties that distinguish them from arbitrary ones" (Moschovakis, 2016a), and seeing as Luzin thought of the two type of equivalence relations as having nice properties, two modern developments with the study of definable equivalence relations attest again to the complexity-regularity interplay.

Theorem 5.6 (Silver (1980))**.** *If E is a $\underset{\sim}{\Pi}^1_1$ equivalence relation on a Polish space, then either E has at most $\aleph_0$ equivalence classes or there exists a perfect set of mutually inequivalent elements.*

Theorem 5.7 (Burgess (1978))**.** *If E is a $\underset{\sim}{\Sigma}^1_1$ equivalence relation on a Polish space, then either E has at most $\aleph_1$ equivalence classes or there exists a perfect set of mutually inequivalent elements.*

The attentive reader will notice that these are instances of the dichotomy theorems considered in Antos and Colyvan (n.d.). Note that these dichotomy theorems are often associated with the topological Vaught's conjecture. We have seen that they have roots that go much further back, to the beginnings of descriptive set theory, and are deeply engaged with classical concerns of regularity and complexity.

5.4 Conclusion and future work

Taking stock, although the emergence of invariant descriptive set theory seems to be the happy accident of two separate strands of developments meeting each other in the 1990s, the present exercise has shown that a number of techniques and concerns taken up by the later Borel equivalence relations theory were actually already present in the early days of descriptive set theory.

The present chapter has only scratched the surface of the pre-history of invariant descriptive set theory, however. There are many more threads to be followed. First and foremost, the developments of the tools that allowed for the abstract study of classification problems and their complexity. This ranges from the discovery of Polish spaces and Polish groups, to the development of the uniformization problems.

Second, the project will benefit from a closer look at the general practice of classification in mathematics. This is a long and complicated development worthy of its own treatment. For example, the use of groups for classification purposes predates everything we have looked at so far by a long margin, according to Kline: "By 1,880 new ideas on groups came into the picture. Klein, influenced by Jordan's work on permutation groups, had shown in his Erlanger Programm that infinite transformation groups, that is, groups with infinitely many elements, could be used to classify geometries." (Kline, 1990)

Finally, we have focused entirely on the mathematical development leading up to the birth of invariant descriptive set theory. The socio-political forces, most notably the Luzin Affair, have been largely ignored. It is not implausible that the Luzin Affair had a significant impact on the development of descriptive set theory, and that it contributed to the emergence of invariant descriptive set theory[24]. I leave this issue wide open for future scholars more well-versed in these matters.

[24]For a parallel development of how socio-political forces shaped the mathematical community in the 1930s, see Parshall (2022).

Bibliography

Aaronson, S. (2014, March 7). *The Scientific Case for $P \neq NP$*. Shtetl-Optimized. Retrieved April 3, 2024, from https://scottaaronson.blog/?p=1720

Aaronson, S. (2016). $P \stackrel{?}{=} NP$. In Jr. Nash John Forbes & M. T. Rassias (Eds.), *Open Problems in Mathematics* (pp. 1–122). Springer International Publishing. https://doi.org/10.1007/978-3-319-32162-2_1

Adleman, L. M., DeMarrais, J., & Huang, M.-D. A. (1997). Quantum Computability. *SIAM Journal on Computing*, *26*(5), 1524–1540. https://doi.org/10.1137/S0097539795293639

Alexandroff, P., & Hopf, H. (1935). *Topologie I*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-02021-0

Antos, C. (2022). Models as Fundamental Entities in Set Theory: A Naturalistic and Practice-based Approach. *Erkenntnis*. https://doi.org/10.1007/s10670-022-00600-3

Antos, C., & Colyvan, M. (n.d.). Explanation in Descriptive Set Theory. In K. Robertson & A. Wilson (Eds.), *Levels of Explanation*. Oxford University Press.

Arana, A. (2009). On Formally Measuring and Eliminating Extraneous Notions in Proofs. *Philosophia Mathematica*, *17*(2), 189–207. https://doi.org/10.1093/philmat/nkn015

Arana, A. (2017). On the Alleged Simplicity of Impure Proof. In R. Kossak & P. Ording (Eds.), *Simplicity: Ideals of Practice in Mathematics and the Arts* (pp. 205–226). Springer International Publishing. https://doi.org/10.1007/978-3-319-53385-8_16

Asperó, D., & Schindler, R. (2021). Martin's Maximum^{++} implies Woodin's axiom $(*)$. *Annals of Mathematics*, *193*(3). https://doi.org/10.4007/annals.2021.193.3.3

Atmai, R. (2015, June 23). *Why does inner model theory need so much descriptive set theory (and vice versa)?* MathOverflow. Retrieved December 22, 2023, from https://mathoverflow.net/q/81939/147031

Avigad, J. (2010, April 19). Gödel and the metamathematical tradition. In S. Feferman, C. Parsons, & S. G. Simpson (Eds.), *Kurt Gödel* (1st ed., pp. 45–60). Cambridge University Press. https://doi.org/10.1017/CBO9780511750762.004

Bagaria, J. (2023). Large cardinals as principles of structural reflection. *Bulletin of Symbolic Logic*, *29*(1), 19–70. https://doi.org/10.1017/bsl.2023.2

Bagaria, J., Castells, N., & Larson, P. (2006). An Ω-logic Primer. In J. Bagaria & S. Todorcevic (Eds.), *Set Theory* (pp. 1–28). Birkh\"auser Basel. https://doi.org/10.1007/3-7643-7692-9_1

Bagaria, J., & Ternullo, C. (2023). *Intrinsic justification for large cardinals and structural reflection.*

Baire, R. (1899). Sur les fonctions de variables réelles. *Annali di Matematica Pura ed Applicata (1898-1922)*, *3*, 1–123.

Baker, A. (2002). Maximizing principles and mathematical methodology. *Logique et Analyse*, *45*(179/180), 269–281.

Barton, N. (2020). Forcing and the Universe of Sets: Must We Lose Insight? *Journal of Philosophical Logic*, *49*(4), 575–612. https://doi.org/10.1007/s10992-019-09530-y

Barton, N., Ternullo, C., & Venturi, G. (2020). On Forms of Justification in Set Theory. *Australasian Journal of Logic*, *17*(4), 158–200. https://doi.org/10.26686/ajl.v17i4.6579

Becker, H., & Kechris, A. S. (1996, December 5). *The Descriptive Set Theory of Polish Group Actions* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511735264

Bernshteyn, A. (2023). Distributed algorithms, the Lovász Local Lemma, and descriptive combinatorics. *Inventiones mathematicae*, *233*(2), 495–542. https://doi.org/10.1007/s00222-023-01188-3

Bernstein, E., & Vazirani, U. (1993). Quantum complexity theory. *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing*, 11–20. https://doi.org/10.1145/167088.167097

Bernstein, E., & Vazirani, U. (1997). Quantum Complexity Theory. *SIAM Journal on Computing*, *26*(5), 1411–1473. https://doi.org/10.1137/S0097539796300921

Berto, F., & Tagliabue, J. (2023). Cellular automata. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Winter 2023). Metaphysics Research Lab, Stanford University.

Boker, U., & Dershowitz, N. (2016). Honest Computability and Complexity. In E. G. Omodeo & A. Policriti (Eds.), *Martin Davis on Computability, Computational Logic, and Mathematical Foundations* (pp. 151–173, Vol. 10). Springer International Publishing. https://doi.org/10.1007/978-3-319-41842-1_6

Borel, É. (1898). *Leçons sur la théorie des fonctions* (Vol. 1). Gauthier-Villars.

Burgess, J., & Miller, D. (1975). Remarks on invariant descriptive set theory. *Fundamenta Mathematicae*, *1*(90), 53–75.

Burgess, J. P. (2008). *Mathematics, models, and modality: Selected philosophical essays*. Cambridge University Press.

OCLC: ocn174130602.

Burgess, J. P. (2015). *Rigor and structure*. Oxford university press.

Cantor, G. (1878). Ein Beitrag zur Mannigfaltigkeitslehre. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, *1878*(84), 242–258. https://doi.org/10.1515/crelle-1878-18788413

Carl, M. (2019, September 23). *Ordinal Computability: An Introduction to Infinitary Machines*. De Gruyter. https://doi.org/10.1515/9783110496154

Cartwright, N. (1983, June 9). *How the Laws of Physics Lie* (1st ed.). Oxford University PressOxford. https://doi.org/10.1093/0198247044.001.0001

Chernavskii, A. V. (2005). On the work of L. V. Keldysh and her seminar. *Russian Mathematical Surveys*, *60*(4), 589–614. https://doi.org/10.1070/RM2005v060n04ABEH003670

Chong, C. T., Nies, A., & Yu, L. (2008). Lowness of higher randomness notions. *Israel Journal of Mathematics*, *166*(1), 39–60. https://doi.org/10.1007/s11856-008-1019-9

Chong, C. T., & Yu, L. (2015). Randomness in the Higher Setting. *The Journal of Symbolic Logic*, *80*(4), 1131–1148. https://doi.org/10.1017/jsl.2015.50

Chong, C. T., Wu, L., & Yu, L. (2019). Basis theorem for Σ^1_2-sets. *The Journal of Symbolic Logic*, *84*(1), 376–387. https://doi.org/10.1017/jsl.2018.81

Church, A. (1936). An Unsolvable Problem of Elementary Number Theory. *American Journal of Mathematics*, *58*(2), 345. https://doi.org/10.2307/2371045

Clarke-Doane, J. (2015). Justification and explanation in mathematics and morality. *Oxford studies in metaethics*, *10*, 80–103.

Cohen, D. E. (1987). *Computability and logic*. E. Horwood ; Halsted Press.

Colyvan, M., Cusbert, J., & McQueen, K. (2018, June 21). Two Flavours of Mathematical Explanation. In A. Reutlinger & J. Saatsi (Eds.), *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780198777946.003.0012

Conley, C., & Miller, B. (2017). Measure reducibility of countable Borel equivalence relations. *Annals of Mathematics*, *185*(2). https://doi.org/10.4007/annals.2017.185.2.1

Copeland, B. J. (2024). The Church-Turing Thesis. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2024). Metaphysics Research Lab, Stanford University.

Crossley, J. N. (2006). Reminiscences of logicians. In *Algebra and logic: Papers from the 1974 summer research institute of the australian mathematical society, monash university, australia* (pp. 1–62). Springer.

Davis, M. (1982). Why Gödel didn't have church's thesis. *Information and Control*, *54*(1-2), 3–24. https://doi.org/10.1016/S0019-9958(82)91226-8

Dawson Jr., J. W. (2006). Why do mathematicians re-prove theorems? *Philosophia Mathematica*, *14*(3), 269–286. https://doi.org/10.1093/philmat/nkl009

Dawson Jr., J. W. (2015). *Why Prove it Again?: Alternative Proofs in Mathematical Practice*. Springer International Publishing. https://doi.org/10.1007/978-3-319-17368-9

De Benedetto, M. (2021). Explication as a Three-Step Procedure: The case of the Church-Turing Thesis. *European Journal for Philosophy of Science*, *11*(1), 21. https://doi.org/10.1007/s13194-020-00337-2

de la Vallée Poussin, C. (1916). *Intégrales de lebesgue. Fonctions d'ensemble. Classes de baire*. Gauthier-Villars.

Delahaye, J.-P. (1993). Randomness, Unpredictability and Absence of Order: The Identification by the Theory of Recursivity of the Mathematical Notion of Random Sequence. In J.-P. Dubucs (Ed.), *Philosophy of Probability* (pp. 145–167). Springer Netherlands. https://doi.org/10.1007/978-94-015-8208-7_8

Demidov, S. S., Levshin, B. V., & Cooke, R. (Eds.). (2016). *The Case of Academician Nikolai Nikolaevich Luzin*. American Mathematical Society.

Detlefsen, M., & Arana, A. (2011). Purity of methods. *Philosophers' Imprint*, *11*.

Di Nasso, M., Goldbring, I., Jin, R., Leth, S., Lupini, M., & Mahlburg, K. (2015). On a Sumset Conjecture of Erdős. *Canadian Journal of Mathematics*, *67*(4), 795–809. https://doi.org/10.4153/CJM-2014-016-0

Di Nasso, M., Goldbring, I., & Lupini, M. (2019). *Nonstandard Methods in Ramsey Theory and Combinatorial Number Theory* (Vol. 2239). Springer International Publishing. https://doi.org/10.1007/978-3-030-17956-4

Downey, R. G., & Hirschfeldt, D. R. (2010, October 29). *Algorithmic Randomness and Complexity*. Springer Science & Business Media.

Duanmu, H., Rosenthal, J. S., & Weiss, W. (2021). *Ergodicity of Markov processes via nonstandard analysis*. American Mathematical Society.

Effros, E. G. (1965). Transformation Groups and C*-algebras. *Annals of Mathematics*, *81*(1), 38–55. https://doi.org/10.2307/1970381

Epstein, R. L., & Carnielli, W. A. (2008). *Computability: Computable functions, logic and the foundations of mathematics* (3. ed). Advanced Reasoning Forum.

Feferman, S. (1962). Transfinite recursive progressions of axiomatic theories. *Journal of Symbolic Logic*, *27*(3), 259–316. https://doi.org/10.2307/2964649

Fenstad, J., & Normann, D. (1974). On absolutely measurable sets. *Fundamenta Mathematicae*, *81*(2), 91–98. https://doi.org/10.4064/fm-81-2-91-98

Ferreirós, J. (2023). The Early Development of Set Theory. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2023). Metaphysics Research Lab, Stanford University.

Feynman, R. P., Leighton, R. B., & Sands, M. L. (1964). *The Feynman lectures on physics* (Vol. 2: Mainly electromagnetism and matter). Pearson Addison Wesley.

Foreman, M. (2010). Models for measure preserving transformations. *Topology and its Applications*, *157*(8), 1404–1414. https://doi.org/10.1016/j.topol.2009.06.021

Foreman, M. (2020). Gödel Diffeomorphisms. *The Bulletin of Symbolic Logic*, *26*(3-4), 219–223. https://doi.org/10.1017/bsl.2020.36

Foreman, M., Rudolph, D., & Weiss, B. (2011). The conjugacy problem in ergodic theory. *Annals of Mathematics*, *173*(3), 1529–1586. https://doi.org/10.4007/annals.2011.173.3.7

Friedman, H. (1981). On the necessary use of abstract set theory. *Advances in Mathematics*, *41*(3), 209–280.

Friedman, H., & Stanley, L. (1989). A Borel Reductibility Theory for Classes of Countable Structures. *Journal of Symbolic Logic*, *54*(3), 894–914. https://doi.org/10.2307/2274750

Friedman, S.-D., Rathjen, M., & Weiermann, A. (2013). Slow consistency. *Annals of Pure and Applied Logic*, *164*(3), 382–393.

Gandy, R. (1988). The confluence of ideas in 1936. *A Half-Century Survey on The Universal Turing Machine*, 55–111.

Gao, S. (2004). The homeomorphism problem for countable topological spaces. *Topology and its Applications*, *139*(1-3), 97–112. https://doi.org/10.1016/j.topol.2003.09.005

Gao, S. (2008, September 3). *Invariant Descriptive Set Theory* (0th ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781584887942

Garey, M. R., & Johnson, D. S. (2009). *Computers and intractability: A guide to the theory of NP-completeness* (27. print). Freeman.

Glasner, E., & King, J. L. (1998). A zero-one law for dynamical properties. *Contemporary Mathematics*, *215*, 231–242.

Glimm, J. (1961). Locally compact transformation groups. *Transactions of the American Mathematical Society*, *101*(1), 124–138. https://doi.org/10.1090/S0002-9947-1961-0136681-X

Gödel, K. (1934, May). On undecidable propositions of formal mathematics systems (S. C. Klein & J. B. Rosser, Eds.). *Institute for Advanced Study.*

Gödel, K. (1986). On the length of proofs. *Kurt Gödel: Collected Works*, *1*, 396–399.

Gödel, K. (2001). *Collected works. 1: Publications 1929 - 1936* (1. issued as paperback). Oxford Univ. Pr.

Goldbring, I., & Hart, B. (2021). Operator algebras with hyperarithmetic theory. *Journal of Logic and Computation*, *31*(2), 612–629. https://doi.org/10.1093/logcom/exaa059

Goldfeld, D. (2004). The Elementary Proof of the Prime Number Theorem: An Historical Perspective. In D. Chudnovsky, G. Chudnovsky, & M. Nathanson (Eds.), *Number Theory* (pp. 179–192). Springer New York. https://doi.org/10.1007/978-1-4419-9060-0_10

Gowers, T. (2007, October 4). *When are two proofs essentially the same?* Gowers's Weblog. Retrieved March 8, 2024, from https://gowers.wordpress.com/2007/10/04/when-are-two-proofs-essentially-the-same/

Greenberg, N. (2020). Two applications of admissible computability. In *Contemporary Logic and Computing* (pp. 604–637). College publications, London.

Greenberg, N. (n.d.). *Proof that there are incomparable Turing degrees*. https://homepages.ecs.vuw.ac.nz/~greenberg/incomparable_degrees.pdf

Hadamard, J. (1905). Cinq lettres sur la théorie des ensembles. *Bulletin de la Société mathématique de France*, *33*, 261–273.

Hallett, M. (1988). *Cantorian Set Theory and Limitation of Size*. Clarendon Press ; Oxford University Press.

Halmos, P. R. (1978). Invariant Subspaces. *The American Mathematical Monthly*, *85*(3), 182–183. https://doi.org/10.1080/00029890.1978.11994549

Halmos, P. R., & Von Neumann, J. (1942). Operator Methods in Classical Mechanics, II. *The Annals of Mathematics*, *43*(2), 332. https://doi.org/10.2307/1968872

Hamkins, J. D. (2016, March 27). *Is there a class of mathematical structures with non-isomorphic natural representations as a standard Borel space?* MathOverflow. Retrieved May 22, 2023, from https://mathoverflow.net/q/234585/147031

Hamkins, J. D. (2020). *Proof and the art of mathematics*. The MIT Press.

Hamkins, J. D. (2022). *Nonlinearity and illfoundedness in the hierarchy of large cardinal consistency strength*.

Hamkins, J. D. (n.d.). *The continuum hypothesis*. Retrieved December 25, 2023, from https://www.infinitelymore.xyz/p/the-continuum-hypothesis

Hamkins, J. D., & Lewis, A. (2000). Infinite time Turing machines. *Journal of Symbolic Logic*, *65*(2), 567–604. https://doi.org/10.2307/2586556

Hamkins, J. D., & Seabold, D. E. (2012). *Well-founded Boolean ultrapowers as large cardinal embeddings*.

Harrington, L. A., Kechris, A. S., & Louveau, A. (1990). A Glimm-Effros Dichotomy for Borel Equivalence Relations. *Journal of the American Mathematical Society*, *3*(4), 903–928. https://doi.org/10.1090/S0894-0347-1990-1057041-5

Hartmanis, J., Chang, R., Chari, S., Ranjan, D., & Rohatgi, P. (1993, June). Relativization: A Revisionistic Retrospective. In *Current Trends in Theoretical Computer Science* (pp. 537–547, Vol. Volume 40). WORLD SCIENTIFIC. https://doi.org/10.1142/9789812794499_0040

Hausdorff, F., & Brieskorn, E. (2008). *Felix hausdorff-gesammelte werke band III: Mengenlehre (1927, 1935) deskripte mengenlehre und topologie* (Vol. 3). Springer Science & Business Media.

Henson, C. W., & Keisler, H. J. (1986). On the strength of nonstandard analysis. *The Journal of Symbolic Logic*, *51*(2), 377–386. https://doi.org/10.2307/2274061

Heron, J. (2021). Set-theoretic justification and the theoretical virtues. *Synthese*, *199*(1-2), 1245–1267.

Hjorth, G. (2000). *Classification and Orbit Equivalence Relations*. American Mathematical Society.

Hjorth, G. (2010). Borel Equivalence Relations. In M. Foreman & A. Kanamori (Eds.), *Handbook of Set Theory* (pp. 297–332). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5764-9_5

Hudson, R. (2020a). The Reality of Jean Perrin's Atoms and Molecules. *The British Journal for the Philosophy of Science*, *71*(1), 33–58. https://doi.org/10.1093/bjps/axx054

Hudson, R. (2020b). What Was Perrin Really Doing in His Proof of the Reality of Atoms? *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, *10*(1), 194–218. https://doi.org/10.1086/707752

Huh, M., Cheung, B., Wang, T., & Isola, P. (2024). The platonic representation hypothesis. *Proceedings of the 41st International Conference on Machine Learning*.

Jech, T. (2003). *Set Theory* (The Third Millennium Edition, revised and expanded). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-44761-X

Kanamori, A. (1995). The Emergence of Descriptive Set Theory. In J. Hintikka (Ed.), *From Dedekind to Gödel: Essays on the development of the foundations of mathematics* (pp. 241–262). Springer Netherlands. https://doi.org/10.1007/978-94-015-8478-4_10

Kanamori, A. (1996). The Mathematical Development of Set Theory from Cantor to Cohen. *The Bulletin of Symbolic Logic*, *2*(1), 1–71. https://doi.org/10.2307/421046

Kanamori, A. (2003). *The Higher Infinite*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-88867-3

Kanamori, A. (2006). Levy and set theory. *Annals of Pure and Applied Logic*, *140*(1-3), 233–252. https://doi.org/10.1016/j.apal.2005.09.009

Kanamori, A., & Magidor, M. (1978). The evolution of large cardinal axioms in set theory. In G. H. Müller & D. S. Scott (Eds.), *Higher Set Theory* (pp. 99–275). Springer Berlin Heidelberg.

Kanovei, V. G. (1988). The axiom of determinacy and the modern development of descriptive set theory. *Journal of Soviet Mathematics*, *40*(3), 257–287. https://doi.org/10.1007/BF01092890

Kanovei, V. G. (1985). The development of the descriptive theory of sets under the influence of the work of Luzin. *Russian Mathematical Surveys*, *40*(3), 135. https://doi.org/10.1070/RM1985v040n03ABEH003591

Kanovei, V. G. (2008). *Borel equivalence relations: Structure and classification*. American Mathematical Society.

OCLC: ocn185095762.

Katz, K. U., & Katz, M. G. (2010). When is .999... less than 1? *The Mathematics Enthusiast*, *7*(1), 3–30. https://doi.org/10.54870/1551-3440.1381

Kaya, B. (2016). *Cantor minimal systems from a descriptive perspective* [Doctoral dissertation].

Kechris, A. S. (2024). *The theory of countable Borel equivalence relations*. Cambridge University Press.

Kechris, A. S. (1999). New Directions in Descriptive Set Theory. *Bulletin of Symbolic Logic*, *5*(2), 161–174. https://doi.org/10.2307/421088

Keldysh, L. V. (1974). The Ideas of N. N. Luzin in Descriptive Set Theory. *Russian Mathematical Surveys*, *29*(5), 179–193. https://doi.org/10.1070/RM1974v029n05ABEH001298

Kennedy, J. (2013). On formalism freeness: Implementing Gödel's 1946 Princeton bicentennial lecture. *Bulletin of Symbolic Logic*, *19*(3), 351–393. https://doi.org/10.2178/bsl.1903030

Kennedy, J. (2020, December 17). *Gödel, Tarski and the Lure of Natural Language: Logical Entanglement, Formalism Freeness* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9780511998393

Kennedy, J., Magidor, M., & Väänänen, J. (2021). Inner models from extended logics: Part 1. *Journal of Mathematical Logic*, *21*(02), 2150012. https://doi.org/10.1142/S0219061321500124

Kleene, S. C. (1981). Origins of Recursive Function Theory. *IEEE Annals of the History of Computing*, *3*(1), 52–67. https://doi.org/10.1109/MAHC.1981.10004

Kleene, S. C. (1952). *Introduction to Metamathematics*. Van Nostrand.

Kline, M. (1990). *Mathematical thought from ancient to modern times. 3*. Oxford University Press.

Koellner, P. (2006). On the Question of Absolute Undecidability†. *Philosophia Mathematica*, *14*(2), 153–188. https://doi.org/10.1093/philmat/nkj009

Koellner, P. (2013). The continuum hypothesis. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Winter 2023). Metaphysics Research Lab, Stanford University.

Koellner, P., & Woodin, W. H. (2009). Incompatible Ω-Complete Theories. *The Journal of Symbolic Logic*, *74*(4), 1155–1170. https://doi.org/10.2178/jsl/1254748685

Koepke, P. (n.d.). *The Influence of Felix Hausdorff on the Early Development of Descriptive Set Theory.*

Koepke, P., & Seyfferth, B. (2009). Ordinal machines and admissible recursion theory. *Annals of Pure and Applied Logic*, *160*(3), 310–318. https://doi.org/10.1016/j.apal.2009.01.005

Kunen, K. (1980). *Set theory: An introduction to independence proofs*. North-Holland.

Kunen, K. (2013). *Set theory* (Rev. ed). College publications, London.

Kuratowski, C., & Tarski, A. (1931). Les opérations logiques et les ensembles projectifs. *Fundamenta Mathematicae*, *17*, 240–248. https://doi.org/10.4064/fm-17-1-240-248

Lebesgue, H. (1905). Sur les fonctions représentables analytiquement. *Journal de mathematiques pures et appliquees*, *1*, 139–216.

Lebesgue, H. (1907). Contribution à l'étude des correspondances de M. Zermelo. *Bulletin de la Societe Mathematique de France*, *35*, 202–212.

Leeuw, K. D., Moore, E. F., Shannon, C. E., & Shapiro, N. (1956, December 31). Computability by Probabilistic Machines. In C. E. Shannon & J. McCarthy (Eds.), *Automata Studies. (AM-34)* (pp. 183–212). Princeton University Press. https://doi.org/10.1515/9781400882618-010

Li, M., & Vitányi, P. (2008). *An Introduction to Kolmogorov Complexity and Its Applications*. Springer New York. https://doi.org/10.1007/978-0-387-49820-1

Loomis, E. S. (1968). *The Pythagorean proposition: Its demonstrations analyzed and classified and bibliography of sources for data of the four kinds of "proofs"*. National Council of Teachers of Mathematics.
OCLC: 432212.

Luzin, N. N. (1927). Sur les ensembles analytiques. *Fundamenta Mathematicae*, *10*, 1–95. https://doi.org/10.4064/fm-10-1-1-95

Luzin, N. N. (1930). *Leçons sur les ensembles analytiques et leurs applications*. Gauthier-Villars.

Luzin, N. N., & Sierpiński, W. (1918). Sur quelques proprietes des ensembles (A). *Bull. Inst. Acad. Sd. Cracovie*, *4*, 35–48.

Maddy, P. (1988a). Believing the axioms. I. *The Journal of Symbolic Logic*, *53*(2), 481–511. https://doi.org/10.2307/2274520

Maddy, P. (1988b). Believing the axioms. II. *The Journal of Symbolic Logic*, *53*(3), 736–764. https://doi.org/10.2307/2274569

Maddy, P. (1997). *Naturalism in Mathematics*. Oxford University Press.
OCLC: 1336405499.

Maddy, P. (1998). V=L and Maximize. In J. A. Makowsky & E. V. Ravve (Eds.), *Logic Colloquium '95* (pp. 134–152, Vol. 11). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-22108-2_10

Maddy, P. (2000). Mathematical Progress. In E. Grosholz & H. Breger (Eds.), *The Growth of Mathematical Knowledge* (pp. 341–352). Springer Netherlands. https://doi.org/10.1007/978-94-015-9558-2_23

Maddy, P. (2007). *Second philosophy: A naturalistic method*. Oxford University Press.
OCLC: ocm76935638.

Maddy, P. (2011, January 27). *Defending the Axioms: On the Philosophical Foundations of Set Theory*. Oxford University Press.

Maddy, P. (2019). What Do We Want a Foundation to Do? In S. Centrone, D. Kant, & D. Sarikaya (Eds.), *Reflections on the Foundations of Mathematics: Univalent Foundations, Set Theory and General Thoughts* (pp. 293–311). Springer International Publishing. https://doi.org/10.1007/978-3-030-15655-8_13

Maddy, P. (2021). Juliette Kennedy. Gödel, Tarski and the Lure of Natural Language: Logical Entanglement, Formalism Freeness. *Philosophia Mathematica*, *29*(3), 428–438. https://doi.org/10.1093/philmat/nkab012

Maddy, P., & Meadows, T. (2020). A reconstruction of Steel's multiverse project. *Bulletin of Symbolic Logic*, *26*(2), 118–169. https://doi.org/10.1017/bsl.2020.5

Mancosu, P., Poggiolesi, F., & Pincock, C. (2023). Mathematical Explanation. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2023). Metaphysics Research Lab, Stanford University.

Marks, A. S., & Unger, S. T. (2017). Borel circle squaring. *Annals of Mathematics*, *186*(2). https://doi.org/10.4007/annals.2017.186.2.4

Meadows, T. (2023a). Beyond Linguistic Interpretation in Theory Comparison. *The Review of Symbolic Logic*, 1–41. https://doi.org/10.1017/S1755020323000321

Meadows, T. (2023b). Risk and theoretical equivalence in mathematical foundations. *Synthese*, *202*(5), 165. https://doi.org/10.1007/s11229-023-04374-1

Miller, B. D. (2009, July 31). *Forceless, ineffective, powerless proofs of descriptive set-theoretic dichotomy theorems*. Sofia, Bulgaria.

Miller, B. D. (2012). The Graph-Theoretic Approach to Descriptive Set Theory. *The Bulletin of Symbolic Logic*, *18*(4), 554–575. https://doi.org/10.2178/bsl.180403

Miller, D. E. (1976). *Invariant descriptive set theory and the topological approach to model theory*. University of California, Berkeley.

Monin, B., & D'Auriac, P.-E. A. (2019). Genericity and Randomness with ITTMs. *The Journal of Symbolic Logic*, *84*(4), 1670–1710. https://doi.org/10.1017/jsl.2019.62

Montalbán, A. (2019). Martin's conjecture: A classification of the naturally occurring Turing degrees. *Notices Amer. Math. Soc*, *66*(8), 1209–1215.

Montalbán, A., & Walsh, J. (2019). On the inevitability of the consistency operator. *The Journal of Symbolic Logic*, *84*(1), 205–225.

Moschovakis, Y. N. (2009). *Descriptive set theory* (2nd ed). American Mathematical Society.

Moschovakis, Y. N. (2016a, January 30). *Effective Descriptive Set Theory: What it is about* (Tutorial). Hejnice, Czech Republic.

Moschovakis, Y. N. (2016b). Hyperarithmetical Sets. In E. G. Omodeo & A. Policriti (Eds.), *Martin Davis on Computability, Computational Logic, and Mathematical Foundations* (pp. 107–149, Vol. 10). Springer International Publishing. https://doi.org/10.1007/978-3-319-41842-1_5

Moschovakis, Y. N. (2018, September 28). *A survey of the origins and development of Descriptive Set Theory.*

Mycielski, J. (1964). Independent sets in topological algebras. *Fundamenta Mathematicae*, *55*(2), 139–147. https://doi.org/10.4064/fm-55-2-139-147

Nies, A. (2012, March 29). *Computability and Randomness*. OUP Oxford.

Novikov, P. S., & Lyapunov, A. M. (1948). Descriptive Set Theory. In *Matematika v SSSR za 30 let (30 years of mathematics in the USSR)* (pp. 243–255). Gostekhizdat.

Nozick, R. (1998). Invariance and Objectivity. *Proceedings and Addresses of the American Philosophical Association*, *72*(2), 21–48. https://doi.org/10.2307/3130880

Odifreddi, P. (1989). *Classical recursion theory: The theory of functions and sets of natural numbers*. North-Holland; Elsevier Science Pub. Co.

Ornstein, D. (1970). Bernoulli shifts with the same entropy are isomorphic. *Advances in Mathematics*, *4*(3), 337–352. https://doi.org/10.1016/0001-8708(70)90029-0

Oxtoby, J. C. (1980). *Measure and Category: A survey of the analogies between topological and measure spaces* (Vol. 2). Springer New York. https://doi.org/10.1007/978-1-4684-9339-9

Pakhomov, F., Rathjen, M., & Rossegger, D. (2024). *Feferman's completeness theorem* (1). https://doi.org/10.48550/ARXIV.2405.09275

Parshall, K. H. (2022). *The new era in American mathematics, 1920-1950*. Princeton University Press.

Perrin, J. (1916). *Atoms* (Dalziel Llewellyn Hammick, Trans.). D. Van Nostrand Company.

Péter, R. (1935). Über den Zusammenhang der verschiedenen Begriffe der rekursiven Funktion. *Mathematische Annalen*, *110*(1), 612–632. https://doi.org/10.1007/BF01448046

Phillips, E. R. (1978). Nicolai Nicolaevich Luzin and the Moscow school of the theory of functions. *Historia Mathematica*, *5*(3), 275–305. https://doi.org/10.1016/0315-0860(78)90114-3

Planck, M. (1993). *A survey of physical theory*. Dover Publications.
Translation of: Physikalische Rundblicke Originally published: A survey of physics. London : Methuen, 1925.

Plunkett, R. L. (1956). A topological proof of a theorem of complex analysis. *Proceedings of the National Academy of Sciences*, *42*(7), 425–426. https://doi.org/10.1073/pnas.42.7.425

Pohlers, W. (1989). *Proof Theory* (Vol. 1407). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-46825-7

Porter, C. (2021). The Equivalence of Definitions of Algorithmic Randomness. *Philosophia Mathematica*, *29*(2), 153–194. https://doi.org/10.1093/philmat/nkaa039

Post, E. L. (1936). Finite combinatory processes—formulation I. *Journal of Symbolic Logic*, *1*(3), 103–105. https://doi.org/10.2307/2269031

Quinon, P. (2021). Can Church's thesis be viewed as a Carnapian explication? *Synthese*, *198*(S5), 1047–1074. https://doi.org/10.1007/s11229-019-02286-7

Raisonnier, J. (1984). A mathematical proof of S. Shelah's theorem on the measure problem and related results. *Israel Journal of Mathematics*, *48*(1), 48–56. https://doi.org/10.1007/BF02760523

Rav, Y. (1999). Why Do We Prove Theorems? *Philosophia Mathematica*, *7*(1), 5–41. https://doi.org/10.1093/philmat/7.1.5

Rescorla, M. (2007). Church's Thesis and the Conceptual Analysis of Computability. *Notre Dame Journal of Formal Logic*, *48*(2). https://doi.org/10.1305/ndjfl/1179323267

Rogers, H. (1959). Computing degrees of unsolvability. *Mathematische Annalen*, *138*(2), 125–140.

Rogers, H. (1987). *Theory of recursive functions and effective computability* (1st MIT Press pbk. ed). MIT Press.

Rudolph, D. (1998). Residuality and orbit equivalence. *Contemporary Mathematics*, *215*, 243–264.

Sabok, M. (2016). Completeness of the isomorphism problem for separable C*-algebras. *Inventiones mathematicae*, *204*(3), 833–868. https://doi.org/10.1007/s00222-015-0625-5

Sacks, G. E. (1966). *Degrees of Unsolvability*. Princeton University Press.

Sacks, G. E. (2016). *Higher recursion theory*. Cambridge University Press.

Salmon, W. C. (2008). *Scientific explanation and the causal structure of the world* (Repr.). Princeton Univ. Pr.

San Mauro, L. (2018). Church-Turing Thesis, in Practice. In M. Piazza & G. Pulcini (Eds.), *Truth, Existence and Explanation: FilMat 2016 Studies in the Philosophy of Mathematics* (pp. 225–248). Springer International Publishing. https://doi.org/10.1007/978-3-319-93342-9_13

Schindler, R. (2011, May 6). *Dilemmas and truths in set theory*. Irvine, California, USA.

Schupbach, J. N. (2018). Robustness Analysis as Explanatory Reasoning. *The British Journal for the Philosophy of Science*, *69*(1), 275–300. https://doi.org/10.1093/bjps/axw008

Sher, G. (2008, September 18). Tarski's Thesis. In D. Patterson (Ed.), *New Essays on Tarski and Philosophy* (pp. 300–339). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199296309.003.0012

Shor, P. (2020, December 15). *Can a mathematical definition be wrong?* MathOverflow. Retrieved March 7, 2024, from https://mathoverflow.net/q/31358/147031

Sieg, W. (2006). Gödel on computability. *Philosophia Mathematica*, *14*(2), 189–207. https://doi.org/10.1093/philmat/nkj005

Sieg, W. (1997). Step by Recursive Step: Church's Analysis of Effective Calculability. *Bulletin of Symbolic Logic*, *3*(2), 154–180. https://doi.org/10.2307/421012

Sierpinski, W. (1918). L'axiome de M. Zermelo et son rôle dans la théorie des ensembles et l'analyse. *Bulletin international de l'Académie des Sciences de Cracovie*, 97–152.

Sierpiński, W. (1917). Sur quelques problèmes qui impliquent des fonctions non mesurables. *Comptes rendus hebdomadaires des seances de l'Academie des sciences*, *164*, 882–884.

Sierpiński, W. (1954). Sur une proposition équivalente à l'existence d'un ensemble de nombres réels de puissance $\aleph_1$. *Bulletin de l'Académie Polonaise des Sciences*, *Classe 3*(2), 53–54.

Sierpiński, W. (1965). *Cardinal and Ordinal Numbers* (2nd ed.). P W N.

Sipser, M. (2013). *Introduction to the theory of computation* (Third edition, international edition). Cengage Learning.

Sorouri, S. (2022). *How would the constructible hierarchy change if you remove parameters?* https://math.stackexchange.com/q/4514711. Mathematics Stack Exchange. https://math.stackexchange.com/q/4514711

Suppes, P. (1988). Philosophical implications of Tarski's work. *The Journal of Symbolic Logic*, *53*(1), 80–91. https://doi.org/10.2307/2274428

Takeuti, G. (1985). Proof theory and set theory. *Synthese*, *62*(2), 255–263. https://doi.org/10.1007/BF00486049

Therrien, V. L. (2020). The Axiom of Choice and the Road Paved by Sierpiński. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, *10*(2), 504–523. https://doi.org/10.1086/710182

Troelstra, A. S. (2011, September 1). History of constructivism in the 20th century. In J. Kennedy & R. Kossak (Eds.), *Set Theory, Arithmetic, and Foundations of Mathematics* (1st ed., pp. 150–179). Cambridge University Press. https://doi.org/10.1017/CBO9780511910616.009

Uspenskii, V. A. (1985). Luzin's contribution to the descriptive theory of sets and functions: Concepts, problems, predictions. *Russian Mathematical Surveys*, *40*(3), 97–134. https://doi.org/10.1070/RM1985v040n03ABEH003590

Viale, M. (2023). Strong forcing axioms and the continuum problem (following Asperó's and Schindler's proof that $\mathbf{MM}^{++}$ implies Woodin's Axiom $(*)$). https://doi.org/10.48550/ARXIV.2305.07784

von Neumann, J. (1932). Zur Operatorenmethode In Der Klassischen Mechanik. *The Annals of Mathematics*, *33*(3), 587. https://doi.org/10.2307/1968537

Wadge, W. W. (1983). *Reducibility and determinateness on the baire space* [Doctoral dissertation, University of California, Berkeley].

Wadge, W. W. (2020, September 7). *Wadge Degrees – the origin story*. Bill Wadge's Blog. Retrieved March 11, 2024, from https://billwadge.com/2020/09/07/wadge-degrees-the-origin-story/

Walsh, J. (2024). On the Hierarchy of Natural Theories. https://doi.org/10.48550/ARXIV.2106.05794

Williams, K. J. (2022, May 18). *Are bi-interpretable theories merely equivalent looks on the same thing? A possible counterexample*. Kameryn Julia Williams. Retrieved July 8, 2024, from http://kamerynjw.net/2022/05/18/bi-interpretability.html

Woodin, W. H. (2010, July 16). *The Axiom of Determinacy, Forcing Axioms, and the Nonstationary Ideal*. De Gruyter. https://doi.org/10.1515/9783110213171